SOFT MATERIALS

SOFT MATERIALS

Structure and Dynamics

Edited by

John R. Dutcher and Alejandro G. Marangoni

University of Guelph
Guelph, Ontario, Canada

MARCEL DEKKER NEW YORK

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress.

ISBN: 0-8247-5358-5

This book is printed on acid-free paper.

Headquarters
Marcel Dekker, 270 Madison Avenue, New York, NY 10016, U.S.A.
tel: 212-696-9000; fax: 212-685-4540

Distribution and Customer Service
Marcel Dekker, Cimarron Road, Monticello, New York 12701, U.S.A.
tel: 800-228-1160; fax: 845-796-1772

World Wide Web
http://www.dekker.com

The publisher offers discounts on this book when ordered in bulk quantities. For more information, write to Special Sales/Professional Marketing at the headquarters address above.

Current printing (last digit):

10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

Preface

Soft materials, such as polymers, biopolymers, liquid crystals, gels, and foams, have physical properties that can be very different from conventional materials, giving rise to intriguing behavior. For example, liquid crystals can behave like either a liquid or a solid, depending on the direction that is chosen within the material. A polymer material can also behave like either a liquid or a solid, depending on the time scale of the measurement. Such unusual behavior occurs commonly for soft materials and can be attributed to several common features that these materials share:

1. The bonding between the molecules can be relatively weak. Because of this, large changes in the materials can be produced by modest changes in their environmental conditions, such as temperature and external fields. The weak bonding also allows the molecules to self-assemble in response to intermolecular interactions to form unique and useful structures over large length scales.
2. The structure of the materials is typically hierarchical, with one level of structure built on underlying levels of structure. In addition, many soft materials are made up of different components, with interfaces between the components playing a large role in determining their physical properties.
3. Their physical properties vary over a large range of time scales. Molecules and assemblies of molecules can move on widely differing time scales from picoseconds—characteristic of the motion of individual segments of macromolecules—to what can be called human time scales (seconds, minutes, years) corresponding to the slow flow or creep of the materials.

The behavior of soft materials is further complicated because they are typically far from equilibrium, with kinetics playing a dominant role in determining their structure and dynamics.

The complex nature of soft materials has made soft materials science a very active area for fundamental research and technological applications. It has become clear that the understanding and control of soft materials requires an interdisciplinary approach, using sophisticated experimental and computational techniques to elucidate the relationships between the structure, dynamics, and function of these materials. In putting together this book, our intent was to bring together contributions from a group of soft materials researchers representing the broad range of academic disciplines involved in this area of research. Because of this, the book is organized into four broad sections, namely, synthetic polymers, complex fluids, biomaterials, and food materials, providing a broad overview of the major areas of scientific and technological endeavor in the field of soft materials science and technology. Moreover, in each section both experimental studies and modeling efforts are presented, so as to provide a more complete picture of a particular research area. It is our hope that researchers working in the field and those who are new to the field will find this book useful.

John R. Dutcher
Alejandro G. Marangoni

Contents

Contributors

Jeff Z. Y. Chen Department of Physics, University of Waterloo, Waterloo, Ontario, Canada

John R. Dutcher Department of Physics, University of Guelph, Guelph, Ontario, Canada

Sharon C. Glotzer Departments of Chemical Engineering and Materials Science and Engineering, University of Michigan, Ann Arbor, Michigan, U.S.A.

H. Douglas Goff Department of Food Science, University of Guelph, Guelph, Ontario, Canada

Yannick Gratton Department of Physics, University of Ottawa, Ottawa, Ontario, Canada

Sarah E. Guthrie Department of Physics and Guelph–Waterloo Physics Institute, University of Waterloo, Waterloo, Ontario, Canada

Jeffrey L. Hutter Department of Physics and Astronomy, The University of Western Ontario, London, Ontario, Canada

Stefan H. J. Idziak Department of Physics and Guelph–Waterloo Physics Institute, University of Waterloo, Waterloo, Ontario, Canada

Martin Kenward Department of Physics, University of Ottawa, Ottawa, Ontario, Canada

Alexandre S. Lemak Ontario Cancer Institute, University of Toronto, Toronto, Ontario, Canada

James R. Lepock Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada

Alejandro G. Marangoni Department of Food Science, University of Guelph, Guelph, Ontario, Canada

Vincent H. Mareau Department of Chemistry, Laval University, Laval, Quebec, Canada

Gianfranco Mazzanti Department of Food Science, University of Guelph, Guelph, Ontario, Canada

Laurette McCormick Department of Physics, University of Ottawa, Ottawa, Ontario, Canada

Rajinder Pal Department of Chemical Engineering, University of Waterloo, Waterloo, Ontario, Canada

David A. Pink Department of Physics, St. Francis Xavier University, Antigonish, Nova Scotia, Canada

Robert E. Prud'homme* Department of Chemistry, Laval University, Laval, Quebec, Canada

Connie B. Roth Department of Physics, University of Guelph, Guelph, Ontario, Canada

Martin G. Scanlon Department of Food Science, University of Manitoba, Winnipeg, Manitoba, Canada

Eric B. Sirota ExxonMobil Research and Engineering Company, Annandale, New Jersey, U.S.A.

Gary W. Slater Department of Physics, University of Ottawa, Ottawa, Ontario, Canada

Francis W. Starr Department of Physics, Wesleyan University, Middletown, Connecticut, U.S.A.

Takuji Tanaka† Department of Food Science, University of Guelph, Guelph, Ontario, Canada

---

Current affiliation:
*Department of Chemistry, University of Montreal, Montreal, Quebec, Canada.
†Department of Applied Microbiology and Food Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

Frédéric Tessier Department of Physics, University of Ottawa, Ottawa, Ontario, Canada

Joost H. J. van Opheusden Mathematics and Statistics Group, Wageningen University, Wageningen, The Netherlands

Rickey Y. Yada Department of Food Science, University of Guelph, Guelph, Ontario, Canada

1

Mobility on Different Length Scales in Thin Polymer Films

CONNIE B. ROTH and JOHN R. DUTCHER University of Guelph, Guelph, Ontario, Canada

I. INTRODUCTION

Thin polymer films with thicknesses of tens of nanometers are studied extensively because they provide an ideal sample geometry for studying the effects of one-dimensional confinement on the structure, morphology, and dynamics of the polymer molecules and because they are used extensively in technological applications such as optical coatings, protective coatings, adhesives, barrier layers, and packaging materials.

Polymer molecules can be prepared in a wide variety of different molecular architectures such as linear molecules, ring molecules, and branched molecules. The simplest molecular architecture corresponds to linear molecules that consist of identical monomer units connected end to end. In a collection of similar molecules, a linear polymer molecule will tend to form in the shape of a random coil. The overall coil size can be characterized statistically by the root-mean-square end-to-end distance R_{ee} which scales as the square root of the length of the molecule or its molecular weight M_w, and typically ranges from several nanometers to tens of nanometers. The polymer molecule can be described by a variety of different length scales, ranging from the size of the individual monomers to the overall size of the molecule R_{ee} (see Fig. 1). The corresponding time scales range from that corresponding to segmental relaxation, related to the glass transition, to that corresponding to the diffusion of entire chains.

In this chapter, we are concerned with the motion of polymer molecules confined to thin films. We begin by describing the basics of the motion of molecules on small length scales, which are related to the glass transition. This will be followed by discussions of mobility of polymer molecules on different length scales and the effects of confinement on molecular motion on different

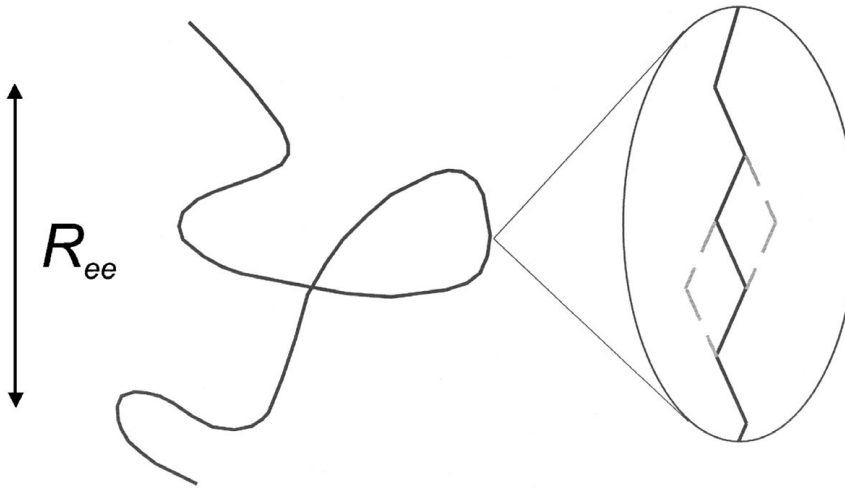


Figure 1 Schematic of a linear polymer molecule, including a magnification of a small number of segments of the molecule. The dashed lines indicate possible motions of the segments of the molecule.

length scales. The remainder of the chapter contains a detailed discussion of the experimental and theoretical studies of the dynamics of thin polymer films. General trends in the data are highlighted and oustanding issues are discussed.

A. The Glass Transition

In the simplest picture, the glass transition describes the transition from a rubberlike liquid to a glassy or amorphous solid as a material is cooled. The glass transition will occur upon cooling for almost any material, from organic liquids to metals to polymers, for a sufficiently fast cooling rate. To measure the glass transition temperature T_g, it is convenient to measure the volume or heat capacity as a function of temperature (see Fig. 2). Typically, one observes an abrupt change in the slope of the temperature dependence of the volume at a certain temperature which is identified as T_g. One also finds that the glass transition temperature measured in such an experiment depends on the cooling or heating rate (see Fig. 2), as well as on the thermal history of the sample. A sample which is cooled faster falls out of equilibrium at a higher temperature, resulting in a T_g value that is higher than that observed for a slower cooling rate.

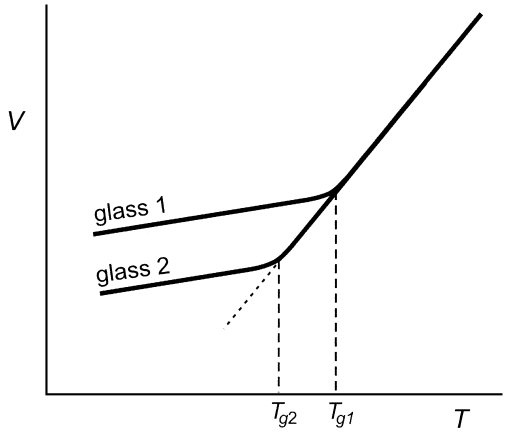


Figure 2 Volume V versus temperature T for a glass-forming material for two different cooling rates. Glass 1 has been cooled faster, resulting in a higher T_g value than that observed for the more slowly cooled glass 2.

Because of the dependence of T_g on time, the glass transition is not a true thermodynamic phase transition, but, rather, it is a kinetic transition in which the motion of the molecules is slowed so dramatically upon cooling that, at sufficiently low temperatures, no appreciable motion of the molecules can occur. The slowing of the dynamics upon cooling can be observed directly by a dramatic increase in the viscosity of the liquid. For example, the viscosity of *o*-terphenyl, a simple glass former, increases by nine orders of magnitude in the 30°C temperature range above T_g [1]. Unlike the large structural changes observed during the formation of a crystal upon cooling, only very subtle structural changes are observed during the formation of a glass, with the structure of the high-temperature liquid effectively frozen in with rapid cooling.

We can understand the frequency (or time) dependence of the measured T_g values if we consider structural relaxation of the material near T_g. If we consider a simplified material that has a single characteristic relaxation time τ at a given temperature, then for short times or correspondingly high frequencies, no significant structural relaxation can occur during the experiment so that the material appears to be solidlike. For times that are much longer than τ or correspondingly low frequencies, the material fully relaxes during the experiment so that the material appears to be liquidlike. If τ increases strongly with decreasing temperature, then below a certain temperature, relaxation for that mobility mode is effectively frozen. For a glass-forming material, phys-

ical quantities such as volume or index of refraction will relax in response to a change in an experimental parameter such as temperature. At a given temperature, the relaxation function $\phi(t)$ can be described by a so-called stretched exponential (Kohlrausch–Williams–Watt or KWW) function:

$$\phi(t) \sim \exp\left[-\left(\frac{t}{\tau}\right)^{\beta}\right] \tag{1}$$

The relaxation function is shown schematically as a plot of ϕ versus t/τ in Fig. 3 for two different values of β. For $\beta = 1$, the decay is a single exponential; for $\beta < 1$, which is typical for glass-forming materials, the decay is nonexponential and is "stretched" in time compared with the exponential decay. The nonexponential behavior has two possible explanations: Either the sample is homogeneous, with each molecule obeying the same nonexponential relaxation, or the sample is dynamically heterogeneous, consisting of regions with different dynamics, each obeying nearly exponential relaxation, with different regions having significantly different relaxation times. For a recent review summarizing the experimental evidence supporting dynamic heterogeneity, see Ref. 2.

As mentioned earlier, cooling of a glass-forming liquid makes it progressively more difficult for the molecules to move significantly on the experimental time scale. This dramatic slowing of the dynamics is difficult to understand theoretically [2,3]. There have been a number of different theo-

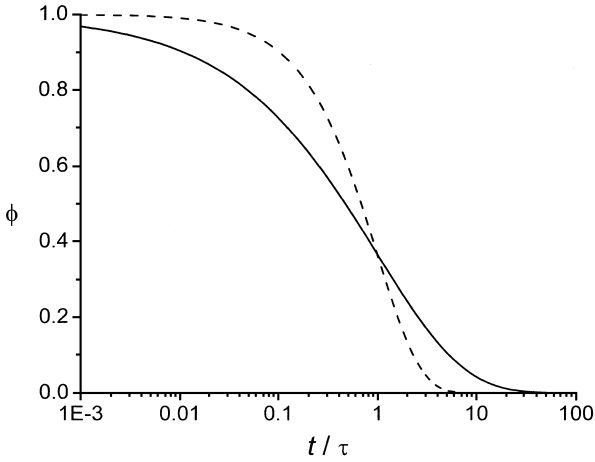


Figure 3 Relaxation function ϕ versus t/τ for an idealized glass-forming material characterized by a single relaxation time τ. The solid line corresponds to $\beta = 0.5$ (stretched exponential) and the dashed line corresponds to $\beta = 1$.

retical approaches to attempt to understand the glass transition, including free-volume, cooperative-motion, and mode-coupling theories. Despite these impressive efforts, no complete theory of the glass transition exists today; for a recent discussion of the relevant issues, see Refs. 3 and 4. However, there are two simple concepts that have been useful in trying to explain the dramatic reduction in molecular mobility with decreasing temperature. The basic ideas of the free-volume and cooperative-motion theories are quite simple. The motion of individual particles in a glass-forming material requires sufficient free volume into which the particles can move. As the temperature is decreased, the density increases and it becomes increasingly difficult for a particle to find sufficient free volume for motion to occur on a reasonable timescale [5]. One way to achieve motion at low temperatures is to allow a cooperative rearrangement of neighboring particles such that many particles must move together if *any* motion is to occur at all. Adam and Gibbs postulated the existence of cooperatively rearranging regions or CRRs as the smallest regions at a given temperature that can rearrange independent of neighboring regions [6]. In their theory, the size of the CRRs was inversely related to the configurational entropy of the system, such that the CRR size increased with decreasing temperature. Experiments indicate that the size of these regions is several nanometers [7,8]. Although the CRRs were originally envisioned as spherical volumes of diameter ξ, recent computer simulations have shown that the cooperative motion is essentially stringlike [9], with the average string length increasing with decreasing temperature. A shortcoming of the Adam–Gibbs approach is the assumption of no interaction between neighboring regions. Because the stringlike shape is less compact than the spherical shape, interaction between neighboring regions will be more significant and this could have a distinctive signature for the effect of confinement on the dynamic properties of glass-forming liquids.

B. Motion of Polymer Molecules on Different Length Scales in Bulk Samples

For polymer molecules, motion can occur on a variety of length scales, from the relaxation of molecular segments, as discussed in the previous section, to diffusion of entire chains. There will be a range of characteristic times at a given temperature corresponding to motion at different length scales, with longer times characterizing motion on larger length scales. If one cools a sample at a certain rate, the mobility modes corresponding to longer relaxation times will freeze out before the mobility modes corresponding to shorter relaxation times. Despite the large difference in the length scale and character of the different motions of a molecule, the temperature dependence of the viscosity, characterizing whole-chain diffusion, is typically very similar

to that for the characteristic times corresponding to the glass transition (α relaxation). Specifically, the temperature dependence of the viscosity and the α-relaxation times is well described by a Vogel–Fulcher function of the form

$$A \exp\left(\frac{T_A}{T - T_0}\right) \tag{2}$$

with similar values of T_A and T_0. In Eq. (2), T_A is an "activation temperature," T_0 is the temperature at which the exponential factor diverges to infinity (typically $T_g - 50$ K for polymers), and A is a temperature-independent prefactor. The similarity of the temperature dependence of dynamic processes that occur on different length scales is a phenomenon known as thermorheological simplicity, and it is the basis for time–temperature superposition which is used extensively in the study of the mechanical properties of bulk polymers [10,11]. A schematic diagram of the time dependence of the compliance $J(t)$ of a bulk linear polymer is shown for two temperatures in Fig. 4 [for a detailed description of $J(t)$, see, e.g., Ref. 10]. The $J(t)$ curve contains regions corresponding to the glass transition (B) and terminal flow or chain diffusion (D) that are separated in time because of the presence of the rubber plateau (C) for high-molecular-weight polymers. Time–temperature superposition means that a change in temperature produces equal shifts in time for the different regions of the $J(t)$ curve. In practice, because the $J(t)$ curve spans many orders of magnitude in time, the complete curve is typically obtained by performing measurements over a limited range of times at

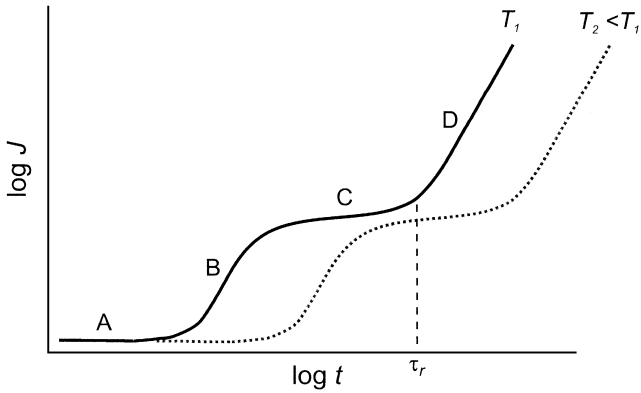


Figure 4 Compliance J versus time t for a bulk linear polymer. Various regions of the compliance curve are labeled. A: glassy plateau: B: glass–rubber transition; C: rubber plateau; D: terminal flow region. The reptation time τ_r for the solid curve is indicated by the vertical dashed line.

different temperatures and shifting the different curves in time using the principle of time–temperature superposition. Although there are deviations from thermorheological simplicity for most polymers, the deviations are typically small and the principle of time–temperature superposition holds reasonably well [12]. It has been suggested that the origin of the similar temperature dependence for motion of polymer molecules at different length scales is related to a property of the individual segments, e.g., segmental frictional force [10].

C. Effects of Confinement on Motion of Polymer Molecules

The idea of observing changes to the dynamics of molecules by confining them to dimensions comparable to that over which motion occurs is appealing. For example, the possible existence of the cooperativity length scale in the Adam–Gibbs theory allows for the possibility of observing so-called finite-size effects in which the size of the sample becomes comparable to the cooperativity length scale as the temperature is varied. Will the mobility of the molecules increase or decrease upon confinement? Imagine a collection of molecules confined to a channel with a diameter d which is on the nanometer length scale. Upon cooling, the cooperativity length scale can become larger than the diameter of the channel. Based on the Adam–Gibbs model, the effect of confinement is to decrease the entropy, which should lead to a slowing down of the dynamics and an increase in T_g. However, the nature of the bounding surface will have a large effect. If there is a large attractive interaction between the molecules and the bounding surface, the mobility will decrease, whereas if the molecules at the bounding surface can move more freely, as is the case of a free or neutral surface, then the overall mobility will increase.

The effect of confinement on molecular mobility has focused either on measurements of simple liquids confined in rather complicated geometries such as porous glasses or measurements of more complex molecules such as polymers confined in simple confining geometries such as thin films. Each type of study can contribute to the overall understanding of the effects of confinement on molecular mobility, but, as will become clear in the discussion below, the results have to be examined carefully to reconcile the measurements with existing and new theories and to resolve apparent contradictions between different measurements on the same and different molecular systems. The picture that emerges is quite coherent, albeit with several surprises and unexplained results.

Finite-size effects on T_g were first observed by Jackson and McKenna for simple liquids confined within porous glasses, which had pore diameters on the nanometer scale [13]. They used differential scanning calorimetry (DSC)

to measure T_g in these systems. Contrary to their initial expectations based on the Adam–Gibbs theory, they observed reductions in T_g, indicating an increase in mobility, as the pore size was decreased. This original work of Jackson and McKenna has led to extensive experimental and theoretical studies of the effects of confinement on the glass transition using a variety of different sample geometries and different molecules (see, e.g., Ref. 14).

By confining polymer molecules to dimensions that are comparable to the different length scales characterizing the molecules, the motion of the molecules can be significantly different than in bulk. Considerable evidence has been obtained that the glass transition temperature T_g of thin polymer films can be substantially different from that in bulk (see, e.g., Refs. 15–17). The effect of confinement on the segmental motion of polymers has been studied using a variety of experimental geometries: interfaces in semicrystalline polymers [18], polymer solutions in porous glasses [19], polymers intercalated into the 2-nm gaps of layered compounds [20], polymer spheres with diameters of tens of nanometers [21], and thin polymer films. We will focus on results obtained for polymer molecules confined to thin films. The simple thin-film geometry is appealing because it is straightforward to control the interactions at the free surface and film–substrate interface, and it is possible to produce films of uniform thickness, and therefore uniform confinement, which can be varied continuously from nanometers to micrometers. In this way, it is possible to probe both chain confinement effects, in which the film thickness is comparable to the overall size of the molecule, as well as finite-size effects.

In contrast to the confinement-induced changes in T_g, it has been found that whole-chain motion in thin polymer films is not substantially different from that in bulk, as discussed in Section II.B. The reconciliation of confinement-induced changes to the segmental mobility with no corresponding changes to whole-chain motion requires the introduction of concepts that are particular to confined systems, such as spatial gradients in molecular mobility near a free surface or interface, or the invalidity of bulk concepts such as time–temperature superposition. In this chapter, we will review recent experimental, theoretical, and computational studies of the effects on the mobility of polymer molecules of confinement to thin films. It is our hope that this critical discussion of the studies to date will lead to a greater understanding of this complex problem.

II. EXPERIMENTAL STUDIES

To probe the molecular mobility in thin polymer films, it is necessary to use a technique that has high sensitivity, because of the small sample volume, or surface sensitivity, which is advantageous because of the large surface-to-

volume ratio in thin films. In general, traditional techniques used to probe the glass transition and whole-chain motion in bulk samples, e.g., DSC and neutron scattering, do not have sufficient sensitivity. It is possible to prepare and measure stacks of identical films using bulk techniques [22], but the requirements for sample preparation are very demanding. New techniques have been developed and existing techniques have been used and adapted to obtain measure of the molecular mobility on different length scales in thin polymer films and at free surfaces of polymers.

A. Glass Transition in Thin Polymer Films

1. Supported Films

The first systematic study of the dependence of the glass transition temperature T_g on film thickness in thin polymer films was performed by Keddie et al. using ellipsometry [23]. In the ellipsometry experiment, the measured change in polarization of light upon reflection from or transmission through the polymer film is used to obtain the thickness h and index of refraction n of the film. To measure the glass transition temperature, the film thickness is measured as a function of temperature and the temperature corresponding to the near discontinuous change in the thermal expansion (see Fig. 2) is identified as T_g. Keddie et al. prepared a series of polystyrene (PS) films on the native oxide layer of silicon wafers with film thicknesses $10 < h < 200$ nm and molecular weights $120 \times 10^3 < M_w < 2900 \times 10^3$. They measured reductions in T_g for $h < 40$ nm, with quantitatively similar results for all M_w values. The data were fitted to an empirical function based on the assumption of the existence of a liquidlike layer at the free surface of the film:

$$T_g(h) = T_g^{\text{bulk}}\left[1 - \left(\frac{\alpha}{h}\right)^{\delta}\right] \tag{3}$$

where $\alpha = 3.2$ nm and $\delta = 1.8$. A compilation of results obtained for PS films on a variety of substrates using numerous experimental techniques such as ellipsometry [23–27], dielectric spectroscopy [28,29], X-ray reflectivity [30], positron annihilation lifetime spectroscopy (PALS) [31], local thermal analysis [32], and probe fluorescence intensity [33–35] is shown in Fig. 5. The trend in the data shown in Fig. 5 is clear: There is a decrease in T_g with decreasing film thickness for all of the data, with good quantitative agreement between data obtained using the different techniques. Because similar results are obtained for PS films on different substrates, one can conclude that the PS films interact only weakly with the underlying substrates. This is not always the case. The importance of the effect of the polymer–substrate interaction on the measured T_g value was first observed for the case of poly(methyl methacrylate) (PMMA) films on Au surfaces and the native silicon oxide

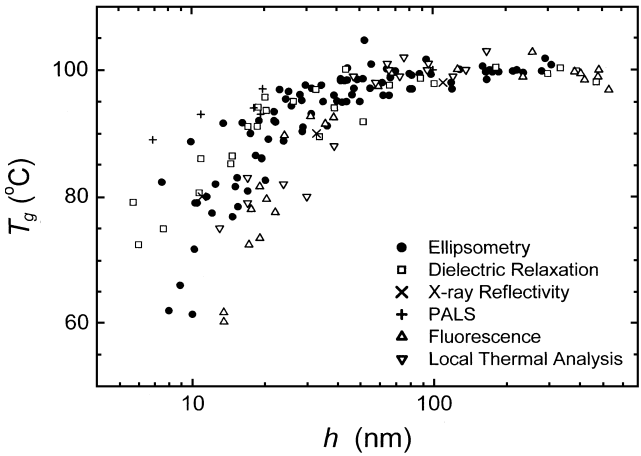


Figure 5 Glass transition temperature T_g as a function of film thickness h measured for PS of different molecular weights M_w supported on a variety of substrates using six different techniques: ellipsometry [23–27], dielectric relaxation [29], X-ray reflectivity [30], PALS [31], fluorescence intensity [35], and local thermal analysis [32]. Note: The data from Refs. 24 and 29 have been shifted in temperature by +3°C so that the bulk T_g values match for all data sets.

layer of silicon wafers. T_g decreased with decreasing film thickness for PMMA films on Au, but a modest increase in T_g was observed with decreasing film thickness for PMMA films on silicon oxide [36]. Subsequently, evidence for a strongly attractive interaction between the polymer and underlying substrate was observed in X-ray reflectivity measurements of poly-(2)-vinylpyridine on a silicon oxide layer on silicon as an increase in T_g with decreasing thickness [37].

As can be seen from the data presented in Fig. 5, there are several different experimental techniques that have been used to measure T_g in supported thin polymer films. All of these techniques probe the material properties averaged across the thickness of the film, yielding an average T_g value for the film. However, it is important to determine whether each of these techniques measures the same physical quantity, e.g., is the T_g value measured by ellipsometry and dielectric relaxation (DR) the same physical quantity? Recent DR measurements of isotactic PMMA sandwiched between Al electrodes [38] and ellipsometry measurements of isotactic PMMA on Al substrates [39] show at least qualitative agreement between the T_g values determined using both techniques. Such comparisons are important for a proper discussion of T_g results obtained using different experimental techniques.

The determination of T_g becomes more difficult as the film thickness is decreased for several reasons: (1) There is a reduction in signal strength because less material is being probed in the experiment; (2) there is a reduction in the contrast between the slopes characterizing the glass and melt regions; and (3) there is a broadening of the transition [27]. Reasons 2 and 3 have been shown to be consistent with the presence of a thin liquidlike layer at the free surface [27]. Despite these effects which make the determination of T_g difficult, T_g values have been obtained for films with thicknesses as small as 6 nm (see Fig. 5).

The presence of the film surfaces creates environments for the polymer segments in contact with the surface that differ from that of segments within the film. In particular, it is reasonable to expect that segments in contact with a free surface are more mobile. Most theoretical models that have been proposed to explain T_g reductions in very thin polymer films begin with this premise and account for variations in segmental mobility across the film thickness as well as novel mechanisms by which the enhanced mobility near the free surface can be transferred deeper into the film [40].

Recently, it has been suggested that attention should be focused on measurements of the distribution of T_g values across the thickness of the film [40,41], rather than measurements of the average T_g value. In response to this suggestion, Ellison and Torkelson have applied the probe fluorescence technique [33,34] to multilayer films incorporating thin layers which contain small quantities of fluorescent probes (either probe molecules or probe-labeled polymer molecules) [35]. The authors have verified that the temperature dependence of the fluorescence intensity provided a measure of the temperature dependence of the density and, therefore, the T_g value of the film and that there is not substantial segregation of the probe molecules to one or both film surfaces. By using the multilayer geometry and varying the position of the fluorescently tagged layer within the film, the authors were able to obtain the distribution of T_g values across the thickness of the film. They found a decrease in T_g near a free surface which extends several tens of nanometers into the film [35]. This large length scale for the effect of the free surface on T_g is considerably larger than that inferred from other measurements on polymer films [27,42] and for other systems using other techniques (see Section IV.B) which have been interpreted in terms of models with layers of different mobility based on measurements of average T_g values. In addition, Ellison and Torkelson have shown that the magnitude of the reduction in T_g at the free surface can also depend on the overall thickness of the film for films with thicknesses less than the spatial extent of the enhanced mobility at a free surface.

Several studies have focused on measuring the distribution of α-relaxation times in thin polymer films, using the techniques of second harmonic

generation (SHG) [43] and DR spectroscopy [28,29,38,44–46]. These studies have allowed a detailed determination of the α-relaxation time distribution in the films. The SHG experiments, performed on a copolymer of isobutyl methacrylate and dye-functionalized methacrylate monomer, revealed no change in the average α-relaxation time with decreasing film thickness, with a corresponding broadening of the α-relaxation time distribution with decreasing film thickness. However, the interaction between the large amounts of dye labels and the substrate may have caused the invariance of the average α-relaxation time with film thickness [34]. DR spectroscopy was used by Fukao et al. to study thin films of PS, poly(vinyl acetate), and atactic PMMA [28,29,44]. For the PS films, the glass transition temperature was identified as the discontinuous change in the temperature dependence of the capacitance, which is determined by the thermal expansion of the polymer film [28]. T_g values determined in this manner decreased with decreasing film thickness, in agreement with the results of other techniques (see Fig. 5). The average α-relaxation temperature T_α was observed to decrease with decreasing film thickness but only below a threshold film thickness value that was small compared with film thicknesses for which T_g reductions were observed using other techniques. In addition, the width of the α-relaxation peak ΔT_α was observed to broaden with decreasing film thickness. Subsequent DR measurements of isotactic PMMA (i-PMMA), which has a large dielectric α-relaxation signal, revealed a decrease in T_g with decreasing film thickness, together with a broadening of the relaxation time distribution and a reduction in the dielectric strength for the α-relaxation process [38]. In this study, the temperature dependence of the relaxation time corresponding to the maximum in the α-relaxation loss peak was fitted to Eq. (2) and T_g was identified as the temperature at which the relaxation time was equal to 100 s. The multilayer film geometry has also been used recently in DR studies of polymer trilayer films consisting of a dielectrically active layer (i-PMMA) sandwiched between two layers of a different polymer (PS) [46]. T_g values for the i-PMMA layers were determined by applying an activation energy fine-structure analysis [47] to the dielectric data. A slight increase in T_g with decreasing i-PMMA film thickness was observed, which is qualitatively different from the decrease in T_g with decreasing film thickness observed for the same i-PMMA films in contact with the metallic electrodes. Dielectric relaxation measurements of i-PMMA films with one free surface have shown that the α process measured at 1 kHz was the same with and without the electrode in contact with the top surface of the polymer film for film thicknesses down to 7 nm [39]. This result is consistent with previous studies that have shown that the presence of thin capping layers on polymer films have no measurable effect on the $T_g(h)$ dynamics [24].

2. Freely Standing Films

The experimental data obtained for polymer films supported on substrates indicate that the presence of the free surface tends to decrease T_g, whereas the presence of the underlying substrate can tend to increase T_g. The T_g value measured for a particular polymer–substrate combination will depend on which film interface dominates. To clarify the role of the free surface on the measured T_g value, the underlying substrate was removed and freely standing polymer films, with air–polymer interfaces on both sides of the film, were studied [24,48–52]. Data points obtained using transmission ellipsometry for freely standing films of narrow distribution PS for different M_w values within the range $575 \times 10^3 < M_w < 9100 \times 10^3$ are indicated with symbols in Fig. 6 [51]. The T_g values obtained using transmission ellipsometry are the same to within experimental uncertainty as the T_g values measured using Brillouin light scattering (BLS) in the original measurements of freely standing PS films [24,48]. This is particularly significant because the two techniques measure different physical properties of the films: BLS measures the light scattering from the rippling of the film surfaces due to thermally excited, film-guided

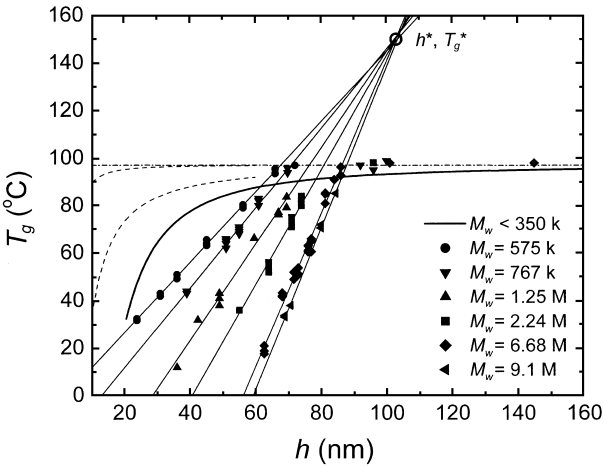


Figure 6 Glass transition temperature T_g as a function of film thickness h for freely standing polystyrene films of different molecular weights. High-M_w data, measured using ellipsometry, are displayed using a different symbol for each M_w value [51], and data for $M_w < 350 \times 10^3$, measured using Brillouin light scattering, are indicated by the thick solid line [50]. The dashed lines for $h < 60$ nm represent the spread and uncertainty in the T_g data obtained for supported PS films (see Fig. 5). The small circle indicates the common intersection point (h^*, T_g^*) of the straight line fits to the reduced T_g data for the high-M_w films.

acoustic phonons and is insensitive to film thickness, whereas the ellipsometry signal is determined by multiple reflections of light from the two film surfaces and yields a measure of the film thickness [53]. Because the BLS signal is determined by the surface ripple, the signal strength does not decrease with decreasing film thickness, unlike all of the other experimental techniques that have been used to probe the glass transition in freely standing polymer films.

The most striking feature of the data presented in Fig. 6 is that the T_g reductions are much larger for freely standing films than they are for supported films. The data obtained for each M_w value has the same qualitative dependence on film thickness h: a constant T_g value equal to that in bulk for sufficiently thick films, and a linear reduction in T_g with h for h less than a threshold film thickness value h_0. This dependence of T_g on h can be interpreted as the interplay between two types of mobility: the bulk mechanism which dominates for sufficiently thick films and a new mode of mobility which becomes more efficient than the bulk mechanism for sufficiently thin films, resulting in reduced T_g values [40]. de Gennes has suggested a mechanism for propagating the mobility of the near-surface segments to depths comparable to the overall size of the polymer molecules, which could give rise to T_g reductions over length scales comparable to the overall size of the polymer molecules [40].

For the $T_g(h)$ data in Fig. 6, there is also a systematic dependence on M_w: an increase in the slope of T_g versus h and the threshold film thickness value h_0 with increasing M_w. The dependence of the $T_g(h)$ data on M_w, or, equivalently, the overall size of the molecules, is unexpected because the glass transition is associated with motion on a much smaller, segmental length scale and also because the $T_g(h)$ data obtained for supported polymer films show no measurable dependence on M_w [23,26]. This so-called chain-confinement effect, in which deviations from bulk behavior are dependent on the overall size of the polymer molecules, is unique to high-M_w freely standing PS films. The freely standing PS film data is even more remarkable if one fits the film thickness dependence of the reduced T_g data for each M_w value and extrapolates the straight line fits to higher temperatures: All six of the straight lines, one for each M_w value, intersect at a single point (h^*, T_g^*) [51,54]. The existence of the intersection point means that, by empirically accounting for the M_w-dependence, all of the reduced T_g values in Fig. 6 can be replotted as a universal scaling plot, as shown in Fig. 7. The collapse of all of the data onto a single line is remarkable. However, the physical significance of the intersection point is not yet understood. To determine if the remarkable scaling behavior shown in Fig. 6 is specific to PS or is more general, measurements have begun on freely standing films of another polymer, PMMA. T_g measurements on freely standing films of atactic PMMA (a-PMMA) of a single M_w value have been reported [52]. A comparison of the $T_g(h)$ data obtained for

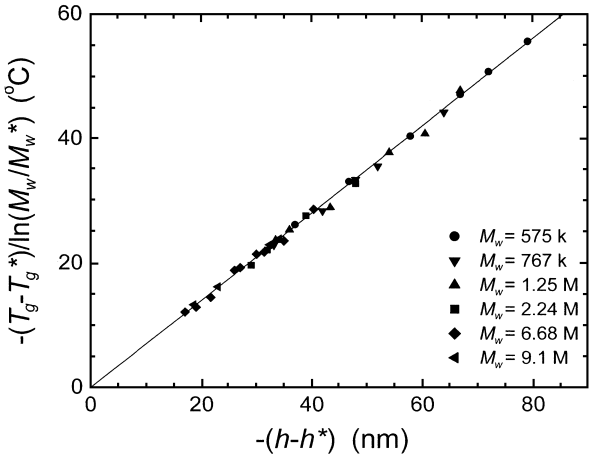


Figure 7 Scaling plot of all of the reduced T_g values for all molecular weights and film thicknesses for the data shown in Fig. 6.

freely standing PS and a-PMMA films of nearly equivalent M_w values is shown in Fig. 8. The same qualitative dependence of T_g on h is observed for both data sets, but the magnitude of the T_g reductions at a given film thickness below the threshold film thickness value is substantially less for the a-PMMA films than for the PS films. This indicates that theoretical models which are proposed to explain T_g reductions in freely standing polymer films will need to account for polymer-specific properties, e.g., chemical structure or steric hindrance.

Measurements performed on films with $M_w < 350 \times 10^3$ (low M_w) have revealed T_g reductions with decreasing film thickness that are comparable in absolute terms to those observed with larger M_w values, but there is no significant M_w dependence of the $T_g(h)$ behavior for the low M_w films (see the solid curve in Fig. 6) [50]. The lack of M_w dependence of the T_g reductions observed for the low-M_w values is reminiscent of the behavior observed for supported polymer films. In fact, the T_g reductions for low-M_w freely standing polymer films are essentially twice as large as those observed for supported polymer films, suggesting that the magnitude of the T_g reduction scales with the number of free surfaces [42]. The low-M_w freely standing polymer film experimental results have been interpreted in terms of a three-layer model in which the layers at the free surfaces are assumed to be more mobile, and the results are consistent with free surface layers which have temperature-dependent thicknesses that scale with temperature in the same manner as the cooperativity length scale [42].

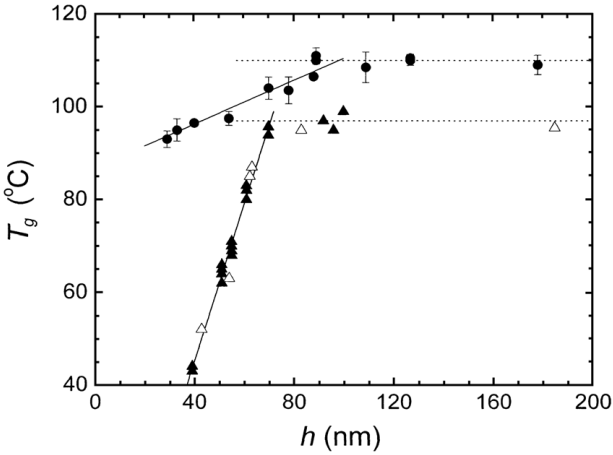


Figure 8 Comparison of $T_g(h)$ results for freely standing PS films (solid triangles: ellipsometry [51]; open triangles: BLS [24]) and PMMA films (solid circles: ellipsometry [52]) of nearly equivalent molecular weights: $M_w = 767 \times 10^3$ for PS and $M_w = 790 \times 10^3$ for a-PMMA.

There are several intriguing aspects to the T_g data obtained for freely standing polymer films. First, the existence of the common intersection point on the plot of T_g versus h for films with different, high-M_w values (see Fig. 6) is striking, but its significance is not yet known. Second, the crossover between a M_w dependence for high-M_w films, indicative of chain-confinement effects, and a lack of M_w dependence for low-M_w films is not understood. Third, the glass transition observed with both BLS and ellipsometry is very sharp (only several degrees Celsius) for all film thicknesses, whereas one would expect that the existence of more mobile surface layers would produce a broadening of the transition for very thin films. All of these aspects of the freely standing polymer film T_g results are intriguing and deserve more careful study.

Recently, T_g measurements have been performed using DSC of aqueous suspensions of PS spheres with diameters of tens of nanometers [21]. It is reasonable to compare these results with those obtained for freely standing polymer films because the spheres in aqueous suspension in the absence of aggregation are in contact only with water and, therefore, are essentially unsupported. By placing many spheres in the suspension, taking care to avoid aggregation, it was possible to obtain sufficient signal using DSC, a technique that is commonly used to measure T_g in bulk samples. In this study, the jump in the specific heat associated with the transition was observed to decrease with decreasing sphere diameter, and these results were interpreted in terms of

a central core with bulk dynamics and an outer shell of 4 nm thickness with faster dynamics. The measured T_g values were not substantially less than the bulk value, in contrast to the large T_g reductions obtained for freely standing PS films. In comparing the results of this study with those obtained for freely standing films, it is important to note two important differences: (1) The PS spheres were produced by microemulsion polymerization in the presence of small quantities of surfactant molecules (to stabilize the spheres in suspension) and therefore have larger polydispersity, different surface character, and less chain orientation than for spin-coated PS freely standing films and (2) the nature of the confinement is different and this has implications for the equilibration of the samples. Obviously, there is a qualitative difference in the confinement of the polymer molecules: Within the spheres, the polymer molecules are confined in three dimensions, instead of the one-dimensional confinement produced in very thin polymer films. The sphere geometry is appealing because it is an equilibrium shape that minimizes the surface area-to-volume ratio. This allows the spheres to be heated above the bulk T_g value without the instabilities that are inherent to thin polymer films (see Section II.B.2).

There have been only preliminary attempts to measure the relaxation dynamics of freely standing PS films [22,55]. A photon correlation spectroscopy (PCS) study revealed a dramatic increase in the relaxation within a 22-nm-thick PS film as the temperature was increased modestly in a small range around room temperature, corresponding to temperatures comparable to the T_g value measured using BLS [55]. The PCS data were well described by a stretched exponential function [Eq. (1)] with a β value that was indistinguishable from that in bulk PS. Inelastic neutron scattering measurements have been performed on stacks of a large number of identical freely standing PS films with a total film thickness of approximately 7.5 μm [22]. Although these measurements were at the edge of feasibility, a small but measurable enhancement of the mean square displacements, and therefore faster dynamics, was observed for a sample consisting of a stack of 140 films, each of 55 nm thickness.

B. Whole-Chain Motion in Thin Polymer Films

The first indication of enhanced mobility in thin polymer films was inferred from optical microscopy studies by Reiter of the breakup or dewetting of thin polymer films at temperatures less than the bulk value of T_g [56]. These measurements necessarily involve motion of entire polymer molecules due to the formation and growth of holes in the films. Subsequent measurements of whole-chain mobility in thin polymer films have focused on diffusion using fluorescently tagged probe and polymer molecules, diffusion at interfaces

between two films, and detailed studies of dewetting and hole formation and growth in thin polymer films. Using these techniques, information has been obtained about whole-chain motion both parallel and perpendicular to the plane of the film.

1. Diffusion in Thin Polymer Films

The first measurements of in-plane chain diffusion were carried out using fluorescence recovery after patterned photobleaching (FRAPP) using fluorescently labeled PS molecules [57]. They found that the in-plane chain diffusion in thin films was substantially slower for film thicknesses as large as 150 nm compared with that measured for very thick films. In subsequent FRAP measurements of in-plane diffusion of rubrene dye molecules in thin PS films at different temperatures, an increase in mobility was observed with decreasing film thickness, but evidence for probe segregation to the film surfaces was also obtained [58]. Although the average concentration of dye molecules in the films is low (0.6–3% by mass), segregation of the dye molecules to the free surface results in much larger concentrations which could artificially enhance the chain mobility by plasticizing the material in the near-surface region.

There have been considerably more studies of diffusion perpendicular to the plane of the film. Fluorescence nonradiative energy transfer (NRET) measurements of films containing layers of acceptor and donor molecules have been performed to measure diffusion perpendicular to the plane of the film, and these measurements have revealed a decrease in the diffusion coefficient for films less than 150 nm thick [59]. Dynamic secondary ion mass spectrometry (SIMS) has been used by Rafailovich and co-workers to measure the diffusion perpendicular to the film plane on a variety of different multilayer film geometries incorporating thin layers of deuterated PS (dPS) and layers of hydrogenated PS (hPS) [60,61] and for freely standing PS films incorporating short dPS chains as probe molecules [62]. The dynamic SIMS measurements have revealed that the diffusion coefficient is either reduced from the bulk value over substantial distances from the substrate and the free surface [60,61] or indistinguishable from that in bulk for freely standing PS films incorporating probe molecules [62]. Neutron reflectometry was used to measure diffusion perpendicular to the film plane for the interface in dPS/hPS bilayer films [63], and it was found that the diffusion coefficient in thin films was unchanged from that in bulk, except for a slight decrease in the diffusion coefficient with the interface placed near the substrate. Kawaguchi et al. have recently used dynamic SIMS and neutron reflectometry to study the interdiffusion of the interface in hPS/dPS bilayer films [64]. They observed interface broadening at temperatures below T_g^{bulk} indicating that there was some enhancement of mobility at the interface compared with the bulk. Boiko

and Prud'homme have used lap-shear strength measurements on PS/PS interfaces which indicate that there is some interdiffusion at temperatures below T_g^{bulk} [65].

How can we reconcile the measurements of diffusion coefficients in thin polymer films, which, in general, show either a decrease or no change with respect to the bulk, with the observation of reduced T_g, which implies enhanced segmental mobility, in similar films? What about comparisons between the diffusion coefficient results for thin polymer films and those obtained from computer simulations of confined polymer molecules, which show increased diffusion parallel to the film plane and decreased diffusion perpendicular to the film plane, as described in Section III? To make meaningful comparisons, it is important to measure the temperature dependence of the mobility at different length scales. In particular, it is important to ask whether or not time–temperature superposition holds for the two types of motion, that is, can the temperature dependence of the mobility at the different length scales be described by Eq. (2) using the same values of T_A and T_0, as it does, to a reasonable approximation, in bulk? Comparisons made at a single temperature could give differences that are due to a thickness dependence of the prefactor A in Eq. (2) instead of a difference in the temperature-dependent factor [15]. To properly compare the mobility at different length scales, it is important to measure the temperature dependence and to compare the values of T_0. Despite the importance of measuring the temperature dependence of the diffusion in thin polymer films, only very few studies to date have provided this information. Clearly, more effort in this area is required to make a proper evaluation of the dynamics at different length scales.

2. Dewetting and Hole Growth

Whole-chain motion in thin polymer films can also be probed by exploiting instabilities that are inherent to thin films. Thin polymer films can be susceptible to the formation of holes when heated to temperatures that are comparable to or greater than the bulk glass transition temperature T_g^{bulk}. Obviously, if holes form and grow in the films, entire polymer molecules must move. This motion occurs initially perpendicular to the film plane, for the hole to form across the entire film thickness, and then is predominantly in the plane of the film, as the hole grows. In the absence of external fields, the instability can be driven by the van der Waals or dispersion interaction, which can be substantial for film thicknesses $h < 100$ nm [66]. For films supported on substrates, it is possible for the dispersion interaction between the two film surfaces to be attractive such that the film of uniform thickness breaks up into droplets via a process known as dewetting [67,68], or repulsive, which enhances the stability of the films [66]. Unsupported or freely standing polymer

films are always unstable to the formation and growth of holes at elevated temperatures [69–71] because the dispersion interaction is always attractive for this film geometry, which is symmetric about the midplane of the film [66]. Holes can form via two different mechanisms: They can be nucleated by external perturbations or defects, such as dust or density inhomogeneities, or they can form spontaneously due to amplification of long-wavelength fluctuations of the film surfaces by the attractive dispersion or van der Waals interaction between the two film surfaces. In the case of nucleation, holes with radii R greater than a critical value R_c grow with time, where $R_c = h/2$ [72]. In the case of spontaneous hole formation, holes can form in the film due to the interplay between the dispersion and surface tension contributions to the free energy [73].

The stresses due to the dispersion interaction and surface tension which drive hole formation and growth are large in very thin films. Therefore, the whole-chain motion in hole formation and growth experiments can be considerably more complicated than that observed in chain diffusion experiments due to the presence of large stresses and the importance of nonlinear viscoelastic effects. Recent studies of the dewetting of supported PS films using scanning probe microscopy (SPM) at temperatures close to T_g^{bulk} have suggested that hole growth may occur, not by polymer flow but by yielding or plastic deformation of the polymer due to the very large stresses that are produced by the interactions that drive the hole growth [74,75]. The importance of nonlinear viscoelastic effects in interpreting these experimental results has also been considered [76–78].

Measurements of hole formation and growth in freely standing polymer films are particularly interesting because of the very large reductions in T_g, indicative of enhanced segmental motion, that have been observed for these films [51]. Hole formation and growth in freely standing films was first measured using optical microscopy [69,71], in which the growth of a single hole was tracked in time. The hole radius was observed to grow exponentially with time: $R(t) = R_0 e^{t/\tau}$, where τ is the characteristic growth time. For freely standing PS films [71], it was found that τ decreased with decreasing film thickness h. This result was interpreted in terms of the bulk phenomenon of shear thinning [79], in which the film viscosity at the edge of the hole, $\eta = \tau\epsilon/h$, where ϵ is the surface tension, decreased with increasing shear-strain rate $\dot{\gamma} = 2/\tau$, according to a power-law dependence $\eta \sim |\dot{\gamma}|^{-d}$, with $d = 0.65 \pm 0.03$. We have recently developed a differential pressure experiment (DPE) to measure the growth of holes in freely standing PS films at temperatures above and below T_g^{bulk} for thicknesses that are sufficiently small such that the T_g values as measured using ellipsometry are reduced from the bulk value [80]. A very small pressure difference (less than 10^{-4} atm) is applied across a freely standing polymer film, and the position of

a piston is controlled to maintain a constant pressure difference across the film. When the film is heated, holes form and grow and the piston must move in one direction to maintain the constant pressure difference. The time dependence of the piston position provides a signature for hole formation and growth and it can be analyzed to obtain the characteristic growth time τ [80]. The pressure difference applied across the film does not influence the formation or growth of the holes; it is chosen to be as small as possible while still allowing the detection of the holes.

The characteristic growth times τ measured at different temperatures using the DPE for freely standing PS films with $M_w = 2440 \times 10^3$ and three different thicknesses are shown in Fig. 9. The T_g values of the freely standing PS films were 97°C (bulk value), 66°C, and 25°C for films with $h = 91$ nm, 68 nm, and 51 nm, respectively [51]. Unfortunately, τ can only be obtained for a limited range of temperatures, because measurements of the characteristic growth time τ are limited by the rapid growth of holes at temperatures $T > T_g^{\text{bulk}}$. The small shifts in temperature between the different data sets shown in Fig. 9 can be explained in terms of the bulk phenomenon of shear thinning [71,80]. The main result of this study is that substantial hole growth occurs only for temperatures that are comparable to T_g^{bulk}, which, for very thin films, can be considerably higher than the reduced T_g values. Thus, the DPE results indicate that despite considerable mobility present on a segmental length

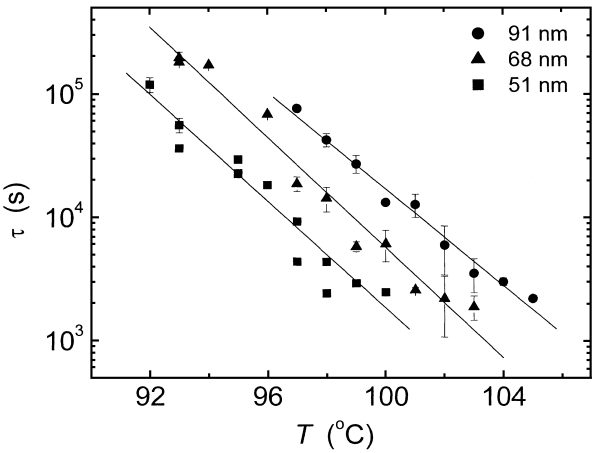


Figure 9 Characteristic growth time τ as a function of temperature T for the growth of holes in freely standing PS films of three different thicknesses with $M_w = 2240 \times 10^3$ as measured using a DPE. (From Ref. 80.)

scale in these films, hole growth and corresponding whole-chain motion does not occur until temperatures close to the bulk value of T_g are reached.

C. Mobility at Surfaces

The large reductions observed in the glass transition temperature in freely standing and supported polymer films suggest the possibility of enhanced mobility near the free surface. Many different techniques have been developed and used to probe the mobility of polymer molecules near a free surface. These experiments usually measure the motion of entire chains near the free surface, instead of segmental mobility, as detailed in the following subsections.

Many measurements have been performed of the response of the free surface to mechanical perturbation using scanning probe microscopy, a variety of techniques have provided noncontact and intermittant-contact measurements of the relaxation of a perturbed surface to infer the mobility of the polymer chains in the near-surface region, and the embedding of small particles in the free surface upon heating has been studied. In addition, PALS has been used as a nonperturbative probe of near-surface segmental mobility.

1. Scanning Probe Microscopy

Scanning probe microscopy (SPM) is a powerful tool for studying the properties of surfaces of a broad range of materials. For viscoelastic materials such as polymers, it is possible not only to measure the surface topography but also the friction and adhesion properties by bringing the SPM tip into contact with the surface and vibrating the tip either vertically or laterally. Measurements are typically performed as a function of temperature, frequency, and the load force on the SPM tip. The first studies [81–83] used lateral tip modulation, and enhanced mobility at the free surface was observed for small molecular weights, which can be understood in terms of segregation of chain ends to the free surface [84]. Further measurements performed using lateral modulation of the SPM tip, also called frictional force or shear modulation force microscopy, have obtained evidence for enhanced mobility at the free surface [85,86], as well as no change with respect to bulk properties [87]. Adhesion measurements have also been performed between polymer surfaces and SPM tips by performing force–distance curve measurements on the surface of polymer films; findings of enhanced surface mobility [88,89] and bulklike properties [90,91] have been reported.

Although it is relatively easy to perform SPM measurements of a free polymer surface at elevated temperatures, the interpretation of these results is not straightforward for a number of reasons. The interaction between a rigid tip and a viscoelastic surface is very complicated [92] and obtaining an accurate estimate of the length scale over which substantial interaction

between the tip and surface occurs is very difficult. The situation is further complicated because the mobility of the polymer chains is expected to vary in the near-surface region. The possibility exists of further complications due to the small contact area between the tip and surface which can result in very large stresses and could lead to nonlinear viscoelastic effects such as shear thinning [93]. It is also important to keep in mind that the results of the SPM experiments depend on the load force and scanning velocity [94]. Finally, it has been pointed out that the inference of T_g values from SPM measurements is difficult because the temperature at which a large tip response is obtained is determined by the Rouse modes and not segmental relaxation [95].

2. Relaxation of a Perturbed Free Surface

Near-surface mobility has been measured by monitoring the smoothing of artificially roughened polymer surfaces by surface tension at elevated temperatures. These measurements have been performed using noncontact or intermittant-contact methods. Tapping-mode SPM has been used to monitor the decrease in surface roughness of PS surfaces that have been produced by preparing the film on a roughened CaF_2 surface [96]. The PS surfaces contained a broad range of roughnesses on different length scales, and the root-mean-square roughness measured over a sample area of $2.5 \times 2.5\ \mu m^2$ was monitored as the samples were heated. They found that partial smoothing of nanometer-sized surface asperities occurred at temperatures below bulk T_g, whereas full relaxation of larger features did not occur until the sample was heated to temperatures above T_g^{bulk} [96]. It has been suggested that the relaxation at smaller length scales could occur at lower temperatures because the increased curvature leads to higher Laplace pressures [75,96], possibly even resulting in plastic deformation of the surface [75]. Measurements of the relaxation of the amplitude of sinusoidal, imprinted surface gratings (period of 600 nm, initial amplitude of 5 nm) on PS surfaces have been performed by diffracting laser light from the grating [93,97–99]. In these measurements, evidence for surface relaxation below the bulk value of T_g was not obtained. The surface grating relaxation measurements have the advantage of probing relaxation of a well-defined wavelength, but observation of the grating amplitude decay by laser light diffraction is limited to wavelengths greater than several hundred nanometers. For smaller-period gratings, tapping-mode SPM could be used to monitor the decay of the amplitude, as in Ref. 96, but it is difficult to produce sinusoidal profiles with periods less than several hundred nanometers.

Rubbing of polymer surfaces with a soft cloth along one in-plane direction orients the polymer chains, and the relaxation of the oriented chains near a free surface at elevated temperatures has been used to obtain a measure of near-surface mobility. In the first study [100], near-edge X-ray absorption fine

structure (NEXAFS) spectroscopy was used to monitor relaxation within the first 1 nm and the first 10 nm from the free surface, and full relaxation was not observed at temperatures below the bulk value of T_g. Recent work using optical birefringence [101,102] and infrared–visible sum frequency generation spectroscopy (SFG) [103] has provided evidence for a surface layer with enhanced mobility that is 12 nm thick. By fitting the orientation factor as a function of temperature to a KWW (stretched exponential) function [see Eq. (1)] and comparing the results for lightly rubbed and highly rubbed surfaces, they found that the activation energy was lower and the β value was larger, indicating a narrower distribution of relaxation times, for the near-surface region [103].

In another approach, NEXAFS has been used to compare the relaxation at different depths near the free surface of a 6-mm-thick PS sample that was uniaxially elongated at room temperature and then heated to 60°C [104]. It was found that the orientation factor of the near-surface region of 2 nm depth decayed with a time constant that was about two-thirds that measured for the near-surface region of 200 nm depth.

It is important to keep in mind that in all of the experiments which measure the relaxation of a deformed polymer surface, the samples are intentionally prepared to be out of equilibrium so that they relax in topography or orientation when heated. To obtain meaningful comparisons between the results of these experiments and those which probe polymer mobility in flat samples that have not been purposely oriented, researchers have been careful to quantify the initial roughness or orientation and to ensure that it is not too large to avoid, e.g., nonlinear relaxation processes. In the studies involving the relaxation of artificially roughened and sinusoidal grating surfaces, care was taken to equilibrate the polymer samples at high temperature while in contact with the roughened surfaces. The roughened polymer surfaces were therefore closer to equilibrium than those measured in the experiments involving rubbed surfaces and uniaxial elongation but, even so, surface-tension-driven roughness on small length scales can lead to plastic deformation of the polymer [75].

3. Embedding of Small Particles in a Free Surface

Surface mobility has also been studied by placing small particles on a polymer surface and measuring the extent to which the particles become embedded in the polymer surface upon heating. Obviously, the deformation of the polymer surface by the particles will depend critically on the wetting behavior of the polymer on the particles. The penetration of small particles into the polymer surface and the adhesion of the particles to polymer surfaces has been measured using the quartz crystal microbalance (QCM) [55,105], X-ray photoelectron spectroscopy [106], and intermittant-contact SPM measure-

ments [107]. The size of the probing particles has been varied from nanometer-sized metal spheres and clusters to micron-sized SiC particles. The results of experiments performed using all of these techniques indicate that embedding occurs at temperatures below that corresponding to bulk T_g. In the case of gold nanoparticles [107], it was inferred that a layer of enhanced mobility that is a few nanometers thick is present at the free surface of PS.

4. Segmental Motion Near a Free Surface

Positron annihilation lifetime spectroscopy provides a near-surface probe, due to limited penetration depth of the positron beam, which probes local mobility. By choosing a low value of the positron energy, greater surface sensitivity is achieved, and mobility was observed to increase as the penetration depth of the positron beam was decreased [108]. The T_g value inferred from the PALS measurement was found to be dramatically reduced by 57°C in the region within 5 nm of the free surface.

III. COMPUTER SIMULATIONS AND THEORETICAL MODELS

In addition to the large body of experimental work studying chain mobility and the glass transition in thin polymer films, there has also been substantial theoretical and computer simulation work on similar systems. This work has been developed both to learn in detail about the structure and dynamics of confined glassy liquids and to explain experimental data obtained for motion on different length scales.

A. Computer Simulations of Confined Polymer Molecules

Ideally, one would like to perform computer simulations on collections of polymer molecules that correspond to the experimental situation, i.e., long polymer chains confined to films at temperatures comparable to T_g, so that meaningful comparisons could be made. However, restrictions on computing power limit these simulations to collections of relatively short, unentangled chains with special computational techniques required either to rapidly quench the samples from high temperature or to achieve relatively low temperatures at more modest quench rates. In general, simulating the dynamics of glassy liquids is very challenging because the timescale for motion becomes extremely slow as the temperature is lowered toward the glass transition. As a result, there have been relatively few simulations of either confined glassy polymer films or supercooled polymer melts, and detailed comparisons between the results that have been obtained from computer simulations and

those obtained from experiments are difficult. Nevertheless, the rich behavior of the static and dynamic properties of confined polymer chains obtained from computational studies provides considerable insight and guidance for the experimental studies.

The computer simulations have focused on the static and dynamic properties of collections of short, unentangled polymer chains confined between walls, and the effect of changing chain length, density, wall separation, wall potential, wall roughness, and temperature. Mansfield and Theodorou [109] used molecular dynamics (MD) to study the static and dynamic properties of freely standing polypropylene films at a temperature lower than the experimental value of T_g. They achieved such low temperatures by using a special equilibration strategy based on consecutive MD runs on different ensembles of the system. They found that the center-of-mass mobility parallel to the walls was enhanced for chains located a distance of the order of R_{ee} from the free surfaces. Baschnagel and Binder [110] used Monte Carlo (MC) techniques to examine the static and dynamic properties of a supercooled polymer melt confined between hard neutral walls. Use of so-called "slithering snake dynamics" [111] allowed the equilibration of the polymer configurations in times that were orders of magnitude faster than those simulated using conventional Rouse dynamics. They found that the motion of the chain segments and the chains parallel to the wall were enhanced and that motion perpendicular to the walls was reduced, over a length scale that was of the order of R_{ee}. Anisotropy in the dynamic properties of confined molecules has also been observed in simulations performed at higher temperatures [112–115]. Recent MD simulations by Baschnagel and co-workers have focused on several features: (1) the modeling of free surfaces using a "softer" wall potential ($\sim z^{-9}$, where z is the distance from the wall) than that within the film ($\sim z^{-12}$) [116], which resulted in enhanced mean square displacements (i.e., accelerated dynamics) next to the walls and (2) the decrease with wall separation of the mode-coupling temperature T_c, instead of T_g, which depends on cooling rate, or the Vogel temperature T_0 [see Eq. (2)], the existence of which is postulated but cannot be derived [116,117]. The effect of the roughness of the walls on the dynamics was studied using MD for supercooled liquids confined to narrow pores [118]. Rough walls produced a large increase in relaxation times close to the wall compared with relaxation times obtained for smooth walls. A hard-sphere MD technique has been used to simulate both supported and freely standing polymer films [119]. An apparent glass transition temperature was identified after a stepped cooling sequence by monitoring the film thickness as a function of temperature and also by extrapolating the temperature dependence of mobility to low temperatures. Slower dynamics in the chains next to the walls and increased T_g values were observed for strongly attractive walls, whereas enhanced dynamics and

substantially decreased T_g values were observed for weakly attractive walls and freely standing films.

B. Theories for Enhanced Mobility in Thin Films

A substantial number of theoretical models have been proposed to explain the observed changes to T_g in thin polymer films. Comparisons between freely standing-film and supported-film T_g results suggest that chain segments near a free surface have a higher mobility than in the bulk. As a result, the models focus on the enhanced mobility at the free surface and mechanisms by which this increased mobility can be propagated deeper into the film. In terms of explaining the large body of experimental results that has been obtained for supported and freely standing films, no one theoretical model is successful in explaining all of the data. Most of the models have been developed specifically to explain the results obtained for supported polymer films and low-M_w freely standing films, for which the T_g reductions are independent of M_w.

A simple scaling argument was proposed to suggest that the T_g value near the free surface will be lower than the bulk value due to segregation of the chain ends to the free surface of the film [84]. This model provides good agreement with experimental results obtained for very low-M_w supported films for which the density of chain ends is substantial [120].

In interpreting the experimental results, various groups have assumed a variation in mobility across the thickness of the film and have modeled this using a layer model in which there is assumed to be a more mobile layer near the free surface and a less mobile layer near the polymer–substrate interface (for supported films). If one assumes that the layer thicknesses are independent of temperature, this leads to a dependence of the difference between the T_g value for the film and the bulk T_g value on the overall film thickness h given by $T_g(h) - T_g^{\text{bulk}} \sim 1/h$. This dependence is certainly not observed experimentally for freely-standing films, and for supported films, for which the T_g reductions are much smaller than for freely standing films, this dependence has not been shown convincingly. Forrest and Mattsson have shown that allowing the mobile layer at the free surface to have a thickness which increases with decreasing temperature, in the same manner as the length scale for cooperative motion, one obtains excellent agreement between the measured T_g values and the average values calculated using the layer model for both low-M_w freely standing films and supported films [42].

A model has been proposed that considers the glass transition obtained upon cooling of thin polymer films as the freezing of the polymer due to memory effects in the viscoelastic modes of the film using concepts from mode-coupling theory [121,122]. Quantitative agreement is achieved with

$T_g(h)$ reductions in low-molecular-weight PS films with only one adjustable parameter, the Young's modulus of the film at T_g.

Ngai has applied the "coupling model" [123–125] to the case of thin polymer films [95,126]. For thin polymer films, it is suggested that the presence of the surface results in orientation of the chain segments parallel to the surface with a corresponding reduction in the number of contacts between segments (intermolecular coupling). The reduction in intermolecular coupling leads to reductions in T_g which are in qualitative agreement with experimental results obtained for supported and freely standing films.

A model based on the concept of percolation incorporating spatial heterogeneity has been proposed in which the film is subdivided into regions with "fast" and "slow" mobility [127,128]. T_g is defined as a three-dimensional percolation of "slow" domains, and reduced values of T_g are obtained for thin films because percolation need only occur over a limited number of layers of the domains (quasi-two-dimensional percolation). This model gives agreement with T_g data obtained for supported polymer films.

One proposed model attributes the $T_g(h)$ reductions observed for thin polymer films to an inhomogeneous density profile, with the $T_g(h)$ value of the inhomogeneous film corresponding to that of a bulk homogeneous sample with the same average density [129]. However, it has been shown that the average density of thin PS films is the same as the bulk density of PS to within 0.5% for freely standing PS films, which have T_g reductions of up to 25% [49], and 1% for supported PS films [130].

One model has been developed specifically to explain the unique M_w-dependent $T_g(h)$ behavior observed for high-M_w freely standing films. Based on free-volume arguments, de Gennes has proposed a new mode of mobility ("sliding mode") to explain $T_g(h)$ reductions in high-M_w freely standing films [40]. As the film thickness is decreased, more chain segments in a given chain contact the free surfaces. The segments in contact with one of the free surfaces are assumed to be more mobile because they have fewer constraints than those segments inside the film. Because of the chain connectivity, the new mode allows for sliding of chain segments that link free-surface segments on the same film surface ("loops") and on different film surfaces ("bridges"). Therefore, this model provides a mechanism by which the enhanced mobility of the segments in contact with the free surfaces can give rise to mobility over distances that are comparable to the overall size of the polymer molecule, and qualitative agreement is obtained with the experimental results for high-M_w freely standing PS films. We point out that because the large T_g reductions observed for freely standing PS films reveal well-defined, reproducible dependences of T_g on h and M_w [51,54], this data set provides a good starting point for critical tests of de Gennes' theory, as well as new theories, of chain confinement effects in the T_g behavior of thin polymer films.

There has also been a very interesting analytical calculation by Semenov for very thin films of long, entangled molecules in which he finds that diffusion of chains is dramatically suppressed due to excluded volume effects which prevent the relaxation of local changes in density due to reptation while segmental mobility is not affected [131]. This mechanism demonstrates that motion on different length scales can be very different indeed.

In this section, we have described many different models that have been developed to interpret the $T_g(h)$ behavior observed experimentally in thin polymer films. Each approach has its successes, but there seems to be no clear reason to choose one approach over another. This situation indicates two difficulties associated with the study of the glass transition in thin polymer films. First, the glass transition in bulk samples is not yet well understood, so the identification of the physical mechanism or mechanisms that are important in confined systems is not clear. Second, the quality and scope of the glass transition data obtained for thin polymer films is not sufficient to allow the preference of one model over the others. Clearly, much additional work, both experimental and theoretical, is required to resolve this issue.

IV. IMPORTANT ISSUES AND QUESTIONS

The experimental, theoretical, and computational studies of thin polymer films that have been performed to date have resulted in considerable understanding of the effect of confinement on polymer mobility, but they have also led to a number of important issues and questions. Some of these issues are directly related to the intriguing results obtained for freely standing films, as discussed earlier: (1) the existence of the common intersection point on the plot of T_g versus h for films with different, large M_w values (see Fig. 6); (2) the crossover between a M_w dependence for high-M_w films, indicative of chain confinement effects, and the lack of M_w dependence for low-M_w films; (3) the sharpness of the glass transition observed with both BLS and ellipsometry for freely standing PS films. All of these features of the freely standing film data deserve more study. In addition, there are more general issues and questions related to mobility in very thin polymer films that impact many of the studies discussed earlier.

A. Mobility on Different Length Scales

Much experimental evidence for the reduction of T_g with decreasing film thickness has been obtained from a variety of studies. In contrast, studies that have focused on measurements of the movement of entire polymer chains have found no measurable enhancement of the chain mobility, and even a decrease in the chain mobility, in very thin polymer films. This difference in

the mobility on different length scales is most dramatic in the case of freely standing films for which very thin films show reductions in T_g of tens of degrees, but no substantial hole formation is observed upon heating until temperatures comparable to the bulk value of T_g are reached [80]. We can suggest several ways in which this apparent contradiction between the results of these two measurements can be resolved. Because the formation of a hole requires the motion of chains across the entire thickness of the film, the dramatic difference between T_g and the temperature at which substantial hole formation and growth occurs upon heating of the films may indicate a variation in mobility across the thickness of the film, with the more mobile segments near the free surface. This is certainly consistent with the existence of a more mobile layer at the free surface of the film inferred from other experiments. The apparent contradiction could also be explained by large differences in the polymer mobility parallel and perpendicular to the film. Formation of a hole requires the motion of entire chains perpendicular to the plane of the film, whereas segmental motion could be enhanced parallel to the plane of the film. This anisotropy in mobility is qualitatively consistent with that observed in computer simulations. The apparent contradiction between the T_g and hole formation data for freely standing PS films could also be an indication of the decoupling of segmental and chain motion in very thin polymer films, as for the mechanism proposed by Semenov [131]. The decoupling of segmental and chain motions in thin polymer films has implications for time–temperature superposition according to Eq. (2), which is found to hold rather well in bulk. However, in thin films, the temperature dependence of the two types of motion might not be the same. This might occur because motions greater than some length scale comparable to the overall size of the molecule are suppressed at temperatures comparable to T_g in very thin polymer films, which is consistent with the experimentally observed decrease in contrast of the thermal expansion of the glass and melt with decreasing film thickness [105]. The current situation is that there are several possible explanations, but no definitive resolution, for this issue, and it highlights the importance of understanding the nature of the mobility being probed in a given experiment so that meaningful comparisons of different experimental studies can be made.

B. Length Scale for Changes in Mobility

A key result in the measurement of T_g in thin polymer films is the observation of changes in the average T_g value with film thickness for film thicknesses that are large compared with the size of the cooperatively rearranging region, which is only several nanometers in size [7,8]. Most theoretical models that have been developed to interpret experimental results have focused on

enhanced mobility near the free surface of a thin polymer film. In addition to models which postulate mechanisms for enhanced mobility over length scales much larger than the change in density near the free surface, considerable success in fitting experimental measurements of the average mobility has been obtained using layer models in which the variation in mobility with distance from the film interfaces is modeled as homogeneous layers of different mobilities. Using layer models, estimates of the spatial extent of the enhanced mobility due to the free surface have been obtained from comparisons between average T_g values measured for PS films of different thickness from PALS measurements [42], heat capacity jumps observed for PS spheres [21], as well as the T_g values inferred for a free surface [108], the relaxation of perturbed surfaces [101–104], and the embedding of small particles [107]. These estimates range from several nanometers to approximately 10 nm. Computer simulations on confined polymer systems, which have the advantage of allowing the determination of detailed information concerning the distance- and direction-dependent mobility with respect to a confining wall, have revealed changes to the dynamics over distances from the confining walls that are comparable to R_{ee} [109,110,116,119], which is of the order of several nanometers for the short molecules used in the simulations.

It is expected that the dynamical behavior will vary smoothly with distance from the free surface. The layer models approximate this smooth variation in a rather crude manner: It is modeled as an abrupt transition from an outer layer of uniformly fast dynamics to an inner layer of uniformly bulk dynamics. Ideally, one would like to have the ability to probe the mobility as a function of distance from the film interfaces, as suggested in Refs. 40 and 41. This has been provided by the recent depth-dependent T_g measurements of Ellison and Torkelson [35]. This detailed study has shown convincingly that there are reductions in T_g near a free surface that extend over distances of tens of nanometers from the free surface. This length scale for enhanced mobility near a free surface is considerably larger than that inferred from layer model interpretations of previous data, and this suggests that the use of layer models may be inadequate to interpret measurements which probe the average mobility of the film. The study of Ellison and Torkelson has also addressed the effect of confinement on the distribution of T_g values with distance from the free surface. In principle, it is possible that T_g reductions near a free surface may be due to the influence of the free surface, confinement within a thin film, or a combination of these effects. Ellison and Torkelson have demonstrated clearly that both free surface and confinement effects are important: There are reductions in T_g near a free surface that extend over distances of tens of nanometers from the free surface, but the magnitude of the reduction at the free surface can also depend on the overall thickness of the film. If the film is sufficiently thin, compared with the spatial extent of T_g

reductions from the free surface, the reduction in T_g at the free surface is less than that observed for the free surface of a very thick film, even though the average T_g value for the thin film is less than that of the thick film. The approach of using fluorescently tagged layers within multilayer films holds much promise to more completely elucidate the detailed T_g behavior near free polymer surfaces and within thin polymer films.

C. Nonequilibrium Nature of Thin Polymer Films

Spin-coating of dilute polymer solutions onto substrates is a powerful technique that allows the deposition of thin polymer films of uniform thickness ranging from nanometers to micrometers. However, the glassy polymer film produced by rapid evaporation of the solvent (solvent quench) differs from that produced by a temperature quench in that there are unwanted solvent molecules present during the solvent quench as well as orientation of the polymer chains along the radial direction. As the solvent content is reduced during spin-coating, during a time determined by the volatility of the solvent, the glassy state "freezes in" below a critical concentration, which for polystyrene in toluene at 25°C is 14% by mass of toluene [132]. This means that the material undergoes a large shrinkage in the glassy state as the remaining solvent molecules are removed. In addition, because evaporation takes place at the free surface of the film, flow instabilities and the formation of a "crust" can be obtained [133–135]. Annealing of the films at temperatures greater than the bulk value of T_g is typically performed after spin-coating to allow the chains to relax, but stress relaxation can be inhibited by attractive interactions between the polymer molecules and the substrate. In addition, there is a tendency for the film to break up or dewet at elevated temperatures due to nucleation of holes at defects, e.g., dust or density inhomogeneities, or by amplification of thermal fluctuations of the free surface of the film for the cases in which the dispersion interaction acting across the film is attractive. Therefore, the annealing process is a trade-off between relaxing the chains and preventing the dewetting of the film, and the best criteria for producing a glassy polymer film are not unambiguously defined.

Ideally, the spin-coated polymer films would be heated to a sufficiently high temperature for a sufficiently long time to allow the molecules to diffuse by a distance which is at least comparable to the overall size of the molecule. This time, also called the reptation time $\tau_r \sim M_w^{3,4}$, can be identified as the time corresponding to the transition between the plateau and terminal flow regions on the $J(t)$ curve shown in Fig. 4. For the annealing of PS films at $T = 115°C$ for 12 h (typical conditions) [51], the annealing time is less than τ_r for $M_w > 850 \times 10^3$. It has been argued that for the purposes of T_g measurements, it is only necessary to anneal the films for times that are sufficient to allow chain

segments between entanglement points to relax, and these times are substantially less than τ_r for high-M_w polymers [51]. One has to also consider the effect on T_g of changes to the chain conformation and entanglement produced by sample preparation, e.g., radial stretching of the molecules due to the spin-coating process. If the T_g reductions were due to the distortion of the chain conformations, then this would be observed for all film thicknesses and the distinctive $T_g(h)$ behavior in Fig. 6, with bulk T_g values for large film thickness and reduced T_g values below a threshold film thickness value, would not be observed. Recent experimental studies have found that reductions in the entanglement density in spin-coated polymer films are not responsible for T_g reductions in thin polymer films [30,136].

V. SUMMARY

It has been 12 years since Jackson and McKenna [13] reported reductions in T_g for organic liquids confined to porous glasses, 10 years since Reiter [56] found evidence for enhanced mobility in dewetting studies of thin polymer films, and nine years since Keddie et al. [23] presented the first systematic study of T_g reductions in thin polymer films with decreasing film thickness. Since then, extensive experimental, theoretical, and computational efforts have focused on trying to explain these original, provocative results. By focusing on well-defined experimental confining geometries such as supported single and multilayer polymer films and freely standing polymer films, with corresponding theories and computer simulations for these geometries, much has been learned of the effect of the film interfaces and free surfaces on the glass transition and whole-chain. These studies have also provided additional surprising results. In particular, the results obtained for freely standing polymer films are unique and remarkable, and further developments of experiments and theories are needed to understand this behavior in detail. In addition, considerable experimental evidence has been obtained that shows that segmental mobility can be enhanced in very thin polymer films and whole-chain motion is unchanged from that in bulk. This behavior is another indication that dynamics in thin polymer films can be considerably different from that in bulk. The development of new experimental techniques to study the dynamics of polymers at free surfaces provides the promise of new insights, but additional effort is required to develop the proper analysis of the data, especially those involving mechanical deformation of polymers at elevated temperatures, so that the results obtained using different techniques can be meaningfully compared. The use of new and improved experimental techniques and sophisticated sample geometries, combined with advances in theoretical analyses and computer simulations, will ensure that the study of dynamics in thin polymer films remains a rich area of research, leading to a

deeper understanding of mobility of confined polymer molecules on different length scales.

ACKNOWLEDGMENTS

We have benefited from stimulating discussions with many researchers in this field that have helped to shape the ideas presented in this review. In particular, we are indebted to Jörg Baschnagel, Kari Dalnoki-Veress, Pierre-Gilles de Gennes, Mark Ediger, Jamie Forrest, Friedrich Kremer, Greg McKenna, Bernie Nickel, and John Torkelson. This work was supported by funding from the PREA program of the Province of Ontario and the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. Brawer, S. *Relaxation in Viscous Liquids and Glasses*; The American Ceramic Society: Columbus, OH, 1985.
2. Ediger, M.D. Annu. Rev. Phys. Chem. 2000, *51*, 99.
3. Donth, E. *The Glass Transition: Relaxation Dynamics in Liquids and Disordered Materials*; Springer-Verlag: Berlin, 2001.
4. Angell, C.A.; Ngai, K.L.; McKenna, G.B.; McMillan, P.F.; Martin, S.W. J. Appl. Phys. 2000, *88*, 3113.
5. Cohen, M.H.; Turnbull, D. J. Chem. Phys. 1959, *31*, 1164.
6. Adam, G.; Gibbs, J.H. J. Chem. Phys. 1965, *43*, 139.
7. Hempel, E.; Hempel, G.; Hensel, A.; Schick, C.; Donth, E. J. Phys. Chem. B 2000, *104*, 2460.
8. Reinsberg, S.A.; Qui, X.H.; Wilhelm, M.; Spiess, H.W.; Ediger, M.D. J. Chem. Phys. 2001, *114*, 7299.
9. Donati, C.; Douglas, J.F.; Kob, W.; Plimpton, S.J.; Poole, P.H.; Glotzer, S.C. Phys. Rev. Lett. 1998, *80*, 2338.
10. Strobl, G. *The Physics of Polymers,* 2nd Ed.; Springer-Verlag: Berlin, 1997.
11. Plazek, D.J.; Zheng, X.D.; Ngai, K.L. Macromolecules 1992, *25*, 4920.
12. Ngai, K.L.; Plazek, D.J. Rubber Chem. Technol. 1995, *68*, 376.
13. Jackson, C.L.; McKenna, G.B. J. Non-Crystal. Solids 1991, *131–133*, 221.
14. Proceedings of the International Workshop on Dynamics in Confinement. J. Phys. IV 2000, *10,* Pr7-3–Pr7-346.
15. Forrest, J.A.; Jones, R.A.L. Polymer Surfaces, In *Interfaces and Thin Films*; Karim, A., Kumar, S., Eds.; World Scientific: Singapore, 2000.
16. Forrest, J.A.; Dalnoki-Veress, K. Adv. Colloid Interf. Sci. 2001, *94*, 167.
17. Special issue on Properties of Thin Polymer Films. Eur. Phys. J E 2002, *8*, 101–266.
18. Schick, M.; Donth, E. Phys. Scr. 1991, *43*, 423.
19. Park, J.-Y.; McKenna, G.B. Phys. Rev. B 2000, *61*, 6667.

20. Anastasiadis, S.H.; Karatasos, K.; Vlachos, G.; Manias, E.; Gianellis, E.P. Phys. Rev. Lett. 2000, *84*, 915.
21. Sasaki, T.; Shimizu, A.; Mourey, T.H.; Thurau, C.T.; Ediger, M.D. J. Chem. Phys. 2003, *119*, 8730.
22. Frick, B.; Dalnoki-Veress, K.; Forrest, J.A.; Dutcher, J.R.; Muray, C.A.; Higgins, A. Eur. Phys. J. E 2003, *12*, S01 022.
23. Keddie, J.L.; Jones, R.A.L.; Cory, R.A. Europhys. Lett. 1994, *27*, 59.
24. Forrest, J.A.; Dalnoki-Veress, K.; Dutcher, J.R. Phys. Rev. E 1997, *56*, 5705.
25. Keddie, J.L.; Jones, R.A.L. Israel J. Chem. 1995, *35*, 21.
26. Tsui, O.K.C.; Zhang, H.F. Macromolecules 2001, *34*, 9139.
27. Kawana, S.; Jones, R.A.L. Phys. Rev. E 2001, *63*, 021501.
28. Fukao, K.; Miyamoto, Y. Phys. Rev. E 2000, *61*, 1743.
29. Fukao, K.; Miyamoto, Y. Europhys. Lett. 1999, *46*, 649.
30. Tsui, O.K.C.; Russell, T.P.; Hawker, C.J. Macromolecules 2001, *34*, 5535.
31. DeMaggio, G.B.; Frieze, W.E.; Gidley, D.W.; Zhu, M.; Hristov, H.A.; Yee, A.F. Phys. Rev. Lett. 1997, *78*, 1524.
32. Fryer, D.S.; Nealey, P.F.; de Pablo, J.J. Macromolecules 2000, *33*, 6439.
33. Ellison, C.J.; Torkelson, J.M. J. Polym. Sci., Part B: Polym. Phys. 2002, *40*, 2745.
34. Ellison, C.J.; Kim, S.D.; Hall, D.B.; Torkelson, J.M. Eur. Phys. J E 2002, *8*, 155.
35. Ellison, C.J.; Torkelson, J.M. Nature Mater. 2003, *2*, 695.
36. Keddie, J.L.; Jones, R.A.L.; Cory, R.A. Faraday Discuss. 1994, *98*, 219.
37. van Zanten, J.H.; Wallace, W.E.; Wu, W.-L. Phys. Rev. E 1996, *53*, R2053.
38. Hartmann, L.; Gorbatschow, W.; Hauwecte, J.; Kremer, F. Eur. Phys. J E 2002, *8*, 145.
39. Sharp, J.S.; Forrest, J.A. Phys. Rev. E 2003, *67*, 031805.
40. de Gennes, P.-G. Eur. Phys. J E 2000, *2*, 201.
41. Jones, R.A.L. Eur. Phys. J E 2000, *2*, 205.
42. Forrest, J.A.; Mattsson, J. Phys. Rev. E 2000, *61*, R53.
43. Hall, D.B.; Hooker, J.C.; Torkelson, J.M. Macromolecules 1997, *30*, 667.
44. Fukao, K.; Uno, S.; Miyamoto, Y.; Hoshino, A.; Miyaji, H. Phys. Rev. E 2001, *64*, 051807.
45. Wübbenhorst, M.; Murray, C.A.; Forrest, J.A.; Dutcher, J.R. Proceedings of the 11th Symposium on Electrets, Melbourne, 2002. New York: IEEE Press.
46. Wübbenhorst, M.; Murray, C.A.; Dutcher, J.R. Eur. Phys. J. E 2003, *12*, S01 025.
47. van Turnhout, J.; Wübbenhorst, M. J. Non-Crystal. Solids 2002, *305*, 50.
48. Forrest, J.A.; Dalnoki-Veress, K.; Stevens, J.R.; Dutcher, J.R. Phys. Rev. Lett. 1996, *77*, 2002; Phys. Rev. Lett. 1996, *77*, 4108.
49. Forrest, J.A.; Dalnoki-Veress, K.; Dutcher, J.R. Phys. Rev. E 1998, *58*, 6109.
50. Mattsson, J.; Forrest, J.A.; Börjesson, L. Phys. Rev. E 2000, *62*, 5187.
51. Dalnoki-Veress, K.; Forrest, J.A.; Murray, C.; Gigault, C.; Dutcher, J.R. Phys. Rev. E 2001, *63*, 031801.

52. Roth, C.B.; Dutcher, J.R. Eur. Phys. J. E 2003, *12*, S01 024.
53. Dutcher, J.R.; Dalnoki-Veress, K.; Forrest, J.A. In *Supramolecular Structure in Confined Geometries*; Warr, G., Manne, S., Eds.; American Chemical Society Symposium Series 1999, *736,* 127–139.
54. Dalnoki-Veress, K.; Forrest, J.A.; de Gennes, P.G.; Dutcher, J.R. J. Phys. IV 2000, *10*, Pr-7–221.
55. Forrest, J.A.; Svanberg, C.; Révész, K.; Rodahl, M.; Torell, L.M.; Kasemo, B. Phys. Rev. E 1998, *58*, R1226.
56. Reiter, G. Europhys. Lett. 1993, *23*, 579.
57. Frank, B.; Gast, A.P.; Russell, T.P.; Brown, H.R.; Hawker, C. Macromolecules 1996, *29*, 6531.
58. Tseng, K.C.; Turro, N.J.; Durning, C.J. Phys. Rev. E 2000, *61*, 1800.
59. Hall, D.B.; Torkelson, J.M. Macromolecules 1998, *31*, 8817.
60. Zheng, X.; Rafailovich, M.H.; Sokolov, J.; Strzhemechny, Y.; Schwarz, S.A.; Sauer, B.B.; Rubinstein, M. Phys. Rev. Lett. 1997, *79*, 241.
61. Pu, Y.; Rafailovich, M.H.; Sokolov, J.; Gersappe, D.; Peterson, T.; Wu, W.-L.; Schwarz, S.A. Phys. Rev. Lett. 2001, *87*, 206101.
62. Pu, Y.; White, H.; Rafailovich, M.H.; Sokolov, J.; Patel, A.; White, C.; Wu, W.-L.; Zaitsev, V.; Schwarz, S.A. Macromolecules 2001, *34*, 8518.
63. Kuhlmann, T.; Kraus, J.; Müller-Buschbaum, P.; Schubert, D.W.; Stamm, M. J. Non-Crystal. Solids 1998, *235–237*, 457.
64. Kawaguchi, D.; Tanaka, K.; Kajiyama, T.; Takahara, A.; Tasaki, S. Macromolecules 2003, *36*, 1235.
65. Boiko, Y.M.; Prud'homme, R.E. J. Polym. Sci., Part B: Polym Phys 1998, *36*, 567.
66. Israelachvili, J.N. *Intermolecular & Surface Forces,* 2nd Ed.; Academic: San Diego, CA, 1991.
67. Reiter, G. Phys. Rev. Lett. 1992, *68*, 75.
68. Brochard-Wyart, F.; Daillant, J. Can. J. Phys. 1990, *68*, 1084.
69. Debrégeas, G.; Martin, P.; Brochard-Wyart, F. Phys. Rev. Lett. 1995, *75*, 3886.
70. Debrégeas, G.; de Gennes, P.-G.; Brochard-Wyart, F. Science 1998, *279*, 1704.
71. Dalnoki-Veress, K.; Nickel, B.G.; Roth, C.; Dutcher, J.R. Phys. Rev. E 1999, *59*, 2153.
72. Gunton, J.D.; San Miguel, M.; Sahni, P.S. In *Phase Transitions and Critical Phenomena*; Domb, C., Lebowitz, J.L., Eds.; Academic Press: London, 1983; Vol. 8.
73. Vrij, A.; Th Hesselink, F.; Lucassen, J.; Van Den Tempel, M. Koninkl. Nederl. Acad. Wetenschappen Amsterdam B 1970, *73*, 124.
74. Reiter, G. Phys. Rev. Lett. 2001, *87*, 186101.
75. Reiter, G. Eur. Phys. J E 2002, *8*, 251.
76. Saulnier, F.; Raphaël, E.; de Gennes, P.-G. Phys. Rev. Lett. 2002, *88*, 196101.
77. Saulnier, F.; Raphaël, E.; de Gennes, P.-G. Phys. Rev. E 2002, *66*, 061607.
78. Shenoy, V.; Sharma, A. Phys. Rev. Lett. 2002, *88*, 236101.
79. Graessley, W.W. Adv. Polym. Sci. 1974, *16*, 1.
80. Roth, C.B.; Nickel, B.G.; Dutcher, J.R.; Dalnoki-Veress, K. Rev. Sci. Instrum. 2003, *74*, 2796.

81. Kajiyama, T.; Tanaka, K.; Satomi, N.; Takahara, A. Macromolecules 1998, *31*, 5150.
82. Satomi, N.; Takahara, A.; Kajiyama, T. Macromolecules 1999, *32*, 4474.
83. Tanaka, K.; Takahara, A.; Kajiyama, T. Macromolecules 2000, *33*, 7588.
84. Mayes, A.M. Macromolecules 1994, *27*, 3114.
85. Hammerschmidt, J.A.; Gladfelter, W.L.; Haugstad, G. Macromolecules 1999, *32*, 3360.
86. Fischer, H. Macromolecules 2002, *35*, 3592.
87. Ge, S.; Pu, Y.; Zhang, W.; Rafailovich, M.; Sokolov, J.; Buenviaje, C.; Buckmaster, R.; Overney, R.M. Phys. Rev. Lett. 2000, *85*, 2340.
88. van Melick, H.; van Dijken, A.; den Toonder, J.; Govaert, L.; Meijer, H. Phil Mag A 2002, *82*, 2093.
89. Bliznyuk, V.N.; Assender, H.E.; Briggs, G.A.D. Macromolecules 2002, *35*, 6613.
90. Tsui, O.K.C.; Wang, X.P.; Ho, J.Y.L.; Ng, T.K.; Xiao, X. Macromolecules 2000, *33*, 4198.
91. Wang, X.P.; Xiao, X.D.; Tsui, O.K.C. Macromolecules 2001, *34*, 4180.
92. Ting, T.C.T. J. Appl. Mech. 1966, *33*, 845.
93. Johannsmann, D. Eur. Phys. J E 2002, *8*, 257.
94. Dinelli, F.; Buenviaje, C.; Overney, R.M. J. Chem. Phys. 2000, *113*, 2043.
95. Ngai, K. Eur. Phys. J E 2002, *8*, 225.
96. Kerle, T.; Lin, Z.; Kim, H.-C.; Russell, T.P. Macromolecules 2001, *34*, 3484.
97. Hamdorf, M.; Johannsmann, D. J. Chem. Phys. 2000, *112*, 4262.
98. Hamdorf, M.; Johannsmann, D. J. Chem. Phys. 2001, *114*, 9685.
99. Petersen, K.; Johannsmann, D. J. Non-Crystal. Solids 2002, *307–310*, 532.
100. Liu, Y.; Russell, T.P.; Samant, M.G.; Stöhr, J.; Brown, H.R.; Cossy-Favre, A.; Diaz, J. Macromolecules 1997, *30*, 7768.
101. Schwab, A.D.; Agra, D.M.G.; Kim, J.H.; Kumar, S.; Dhinojwala, A. Macromolecules 2000, *33*, 4903.
102. Agra, D.M.G.; Schwab, A.D.; Kim, J.H.; Kumar, S.; Dhinojwala, A. Europhys. Lett. 2000, *51*, 655.
103. Schwab, A.D.; Dhinojwala, A. Phys. Rev. E 2003, *67*, 021802.
104. Wallace, W.E.; Fischer, D.A.; Efimenko, K.; Wu, W.-L.; Genzer, J. Macromolecules 2001, *34*, 5081.
105. Forrest, J.A.; Mattsson, J.; Börjesson, L. Eur. Phys. J E 2002, *8*, 129.
106. Zaporojtchenko, V.; Strunskus, T.; Erichsen, J.; Faupel, F. Macromolecules 2001, *34*, 1125.
107. Teichroeb, J.H.; Forrest, J.A. Phys. Rev. Lett. 2003, *91*, 016104.
108. Jean, Y.C.; Zhang, R.; Cao, H.; Yuan, J.-P.; Huang, C.-M.; Nielsen, B.; Asoka-Kumar, P. Phys. Rev. B 1997, *56*, R8459.
109. Mansfield, K.F.; Theodorou, D.N. Macromolecules 1991, *24*, 6283.
110. Baschnagel, J.; Binder, K. J. Phys. I (France) 1996, *6*, 1271.
111. Kremer, K.; Binder, K. Comput. Phys. Rep. 1988, *7*, 259.
112. Mansfield, K.F.; Theodorou, D.N. Macromolecules 1989, *22*, 3143.
113. Bitsanis, I.; Hadziioannau, G. J. Chem. Phys. 1990, *92*, 3827.
114. Matsuda, T.; Smith, G.D.; Winkler, R.G.; Yoon, D.Y. Macromolecules 1995, *28*, 165.

115. Doruker, P.; Mattice, W.L. Macromolecules 1999, *32*, 194.
116. Varnik, F.; Baschnagel, J.; Binder, K. Phys. Rev. E 2002, *65*, 021507.
117. Varnik, F.; Baschnagel, J.; Binder, K. Eur. Phys. J E 2002, *8*, 175.
118. Scheidler, P.; Kob, W.; Binder, K. Europhys. Lett. 2002, *59*, 701.
119. Torres, J.A.; Nealey, P.F.; de Pablo, J.J. Phys. Rev. Lett. 2000, *85*, 3221.
120. Tanaka, K.; Taura, A.; Ge, S.-R.; Takahara, A.; Kajiyama, T. Macromolecules 1996, *29*, 3040.
121. Herminghaus, S.; Jacobs, K.; Seemann, R. Eur. Phys. J E 2001, *5*, 531.
122. Herminghaus, S. Eur. Phys. J E 2002, *8*, 237.
123. Ngai, K.L. Commun. Solid State Phys. 1979, *9*, 121.
124. Tsang, K.-Y.; Ngai, K.L. Phys. Rev. E 1996, *54*, R3067.
125. Tsang, K.-Y.; Ngai, K.L. Phys. Rev. E 1997, *56*, R17.
126. Ngai, K.; Rizos, A.K. Mater. Res. Soc. Symp. Proc. 1997, *455*, 147.
127. Long, D.; Lequeux, F. Eur. Phys. J E 2001, *4*, 371.
128. Merabia, S.; Long, D. Eur. Phys. J E 2002, *9*, 195.
129. McCoy, J.D.; Curro, J.G. J. Chem. Phys. 2002, *116*, 9154.
130. Wallace, W.E.; Beck Tan, N.C.; Wu, W.L.; Satija, S. J. Chem. Phys. 1998, *108*, 3798.
131. Semenov, A.N. Phys. Rev. Lett. 1998, *80*, 1908.
132. McKenna, G.B. J. Phys. IV 2000, *10*, Pr-7–53.
133. Strawhecker, K.E.; Kumar, S.K.; Douglas, J.F.; Karim, A. Macromolecules 2001, *34*, 4669.
134. de Gennes, P.-G. Eur. Phys. J E 2001, *6*, 421.
135. de Gennes, P.-G. Eur. Phys. J E 2002, *7*, 31.
136. Bernazzani, P.; Simon, S.L.; Plazek, D.J.; Ngai, K.L. Eur. Phys. J E 2002, *8*, 201.

2

Crystallization of Thin Polymer Films: Crystallinity, Kinetics, and Morphology

VINCENT H. MAREAU and ROBERT E. PRUD'HOMME * Laval University, Laval, Quebec, Canada

I. POLYMER CRYSTALLIZATION

The mechanisms involved in the crystallization of polymer molecules, in the form of folded chain crystals [1] called lamellae, are now well understood [2–4]. The lamellar organization of polymers, crystallized both from solution and from the melt, is generally observed when the polymer chains are flexible and linear enough to fold back and forth in a lamella much thinner than the contour length of the polymer chain. When polymers are crystallized from dilute solutions at low supercoolings (high temperatures), single crystals are usually formed [2]. In optimal conditions, they are monolayered crystals with lateral dimensions in the range of 1–50 μm and a thickness between 5 and 15 nm. These chain-folded lamellar crystals usually have a geometric shape reflecting the symmetry of the polymer unit cell. Lamellae can also develop into fibrils and spherulites [2–4] when the polymer is isothermally crystallized from the melt.

Morphological observations of spherulitic crystallization were first carried out by optical microscopy (Fig. 1), indicating the presence of radial crystalline fibrils that branch to fill space [5]. The lamellar structure of those fibrils was later investigated by electron microscopy after etching, which allows digging into the spherulite [6,7]. In the center of the spherulite, lamellae were shown to be stacked parallel, but they splay apart and branch when moving away from the center, giving a sheaflike organization that finally develops into a spheru-

* *Current affiliation*: University of Montreal, Montreal, Quebec, Canada.

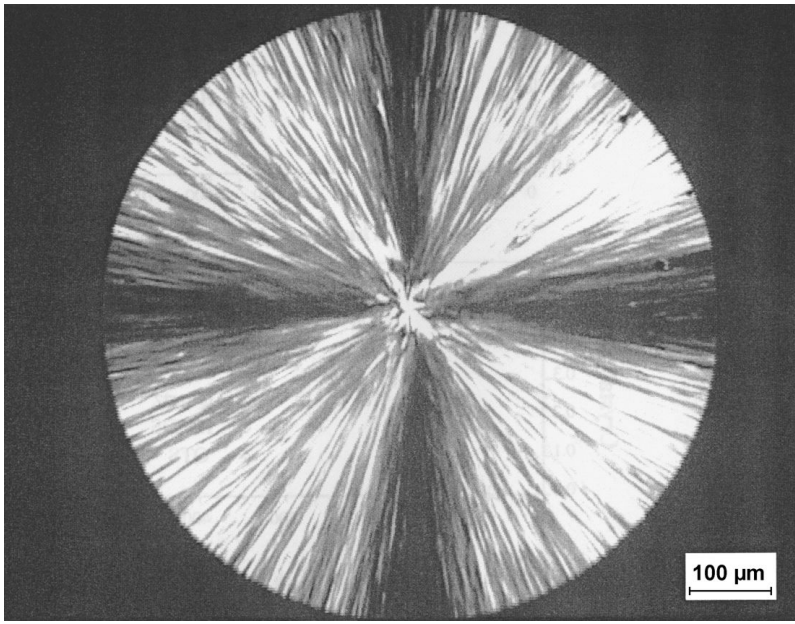


Figure 1 Crossed polar optical microphotograph of a poly(ethylene oxide) spherulite observed during the crystallization of a 10-µm-thick film. The spherulite displays the classical Maltese Cross pattern with radial crystalline fibrils and is surrounded by the melt. (Microphotograph by V. H. Mareau.)

lite when the lamellae grow in every direction of space [8]. Several theories were developed to explain the branching and splaying of the lamellae [9].

Simultaneously, crystallization kinetics measurements were carried out to understand the mechanisms and parameters that control the growth rate of folded-chain crystals [10,11]. Spherulites were found to grow linearly with time upon isothermal crystallization. Growth rates can be represented as a function of crystallization temperature by a bell-shaped curve with a maximum at a temperature intermediate between the glass transition and the melting temperature of the polymer. The secondary nucleation theory of Hoffman and Lauritzen [12] successfully describes this type of phenomenon.

All of the above-described morphological and crystallization studies were done on solution-grown single crystals or on polymer films usually more than 10 µm thick. However, the behavior of thin polymer films has become increasingly important for microelectronic applications that require them. Recent measurements have allowed to study the crystallization of thin (100–1000 nm) and ultrathin (below 100 nm) films, thanks, mainly, to the development of the tapping-mode atomic force microscope (TM-AFM),

because of its ability to obtain *in situ* real-space information on submicron length scales, during the process of crystallization, when used in conjunction with an appropriate hot stage, without significantly damaging the sample [13]. It is the purpose of this chapter to review our present knowledge in this area.

We have divided this review as follows. In Section II, the influence of the film thickness on both crystallinity and spherulitic growth rates in thin and ultrathin polymer films will be described, with specific examples of hindered crystallizations, and decrease of the growth rate when the film thickness is decreased. In Section III, the specific morphologies observed in thin and ultrathin films will be described with, for example, preferential edge-on lamellae orientation, development of single-crystal structures, and dendritic morphologies. In Section IV, we will show how studies involving thin and ultrathin films can become a tool for the understanding of the crystallization of polymers, as they allow a better visualization of the crystallization process. In Section V, the crystallization of miscible poly(ε-caprolactone)/poly(vinyl chloride) blends thin films will be described, highlighting the importance of surface interactions, including those of the surface layer in thick films. In Section VI, recent results involving block copolymers, which are used as templates to confine the polymer crystallization, will be briefly described. The final section will summarize some of the key ideas of this review and outline some anticipated research avenues in this area.

II. CRYSTALLINITY AND GROWTH RATES OF POLYMER THIN AND ULTRATHIN FILMS

As early as 1980, Prest and Luca [14] reported a preferential alignment of the polymer chains in the plane of 1–5-μm-thick polymer films prepared by solvent casting. More than 15 years later, as polymer thin and ultrathin films began to be commonly used in technological applications [15], evidence of the impact of the film thickness on the crystallinity and rate of crystallization of poly(di-*n*-hexylsilane) films were established [16,17].

Despotopoulou et al. studied a series of poly(di-*n*-hexylsilane) films with thicknesses between 5 and 350 nm and reported a disorder of the polymer chains and a reduction of the degree of crystallinity [16] and rate of crystallization [17] for films thinner than 50 nm. The overall crystallinity of the films was measured from ultraviolet (UV) absorption spectra as the crystalline and amorphous phases generate two distinct absorption peaks. Figure 2 gives the degree of crystallinity as a function of film thickness, with three different substrates: hydrophilic quartz and hydrophobic quartz silylated with octadecyltrichlorosilane (OTS) and hydrophobic quartz silylated with 1,1,1,3,3,3-hezamethyldisilazane (HMDS). The degree of crystallinity drops for thicknesses below 50 nm, and the crystallization is nearly totally hindered below 15 nm, both at hydrophilic and hydrophobic surfaces. No

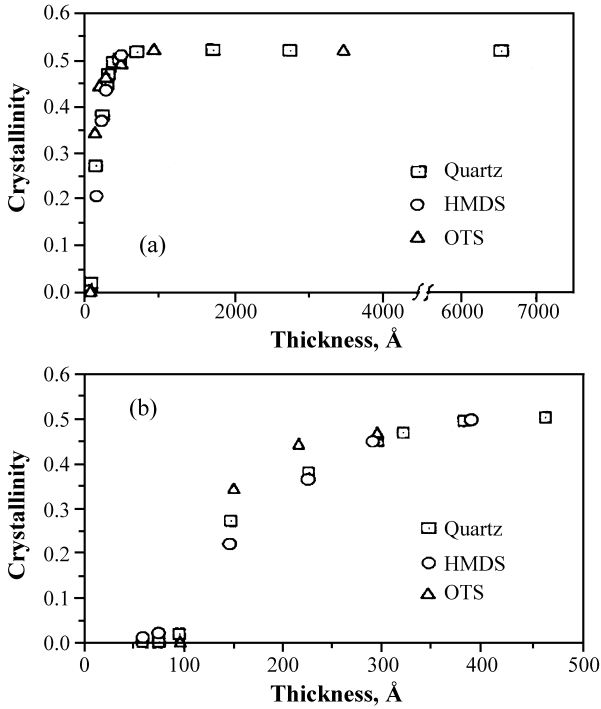


Figure 2 (a) Degree of crystallinity of poly(di-*n*-hexylsilane) films as a function of film thickness on three different substrates: hydrophilic quartz and hydrophobic quartz silylated with OTS or HMDS; (b) magnification of the ultrathin film region highlighting a threshold value of about 15 nm. The crystallization was carried out at 3–5°C for at least 8 h. (Reprinted with permission from Ref. 16. Copyright 1996, John Wiley & Sons, Inc.)

difference was seen between low- and high-molecular-weight polymers, except a higher final degree of crystallinity in the former case. It was concluded that the overall chain length or radius of gyration (coil dimensions) is not the determining factor explaining the decrease in crystallinity. Rather, it was proposed that as the film thickness decreases, nuclei of critical dimensions may have difficulties to develop, even being suppressed when films are thinner than 15 nm.

Kinetics of crystallization was also followed by UV spectroscopy [17], and the overall degree of crystallinity was reported as a function of time for ultrathin films of poly(di-*n*-hexylsilane) crystallized at 5°C (Fig. 3). It was then found that thinner films evolve more slowly than thicker ones. For example, under these conditions, a plateau is reached for the 50-nm sample

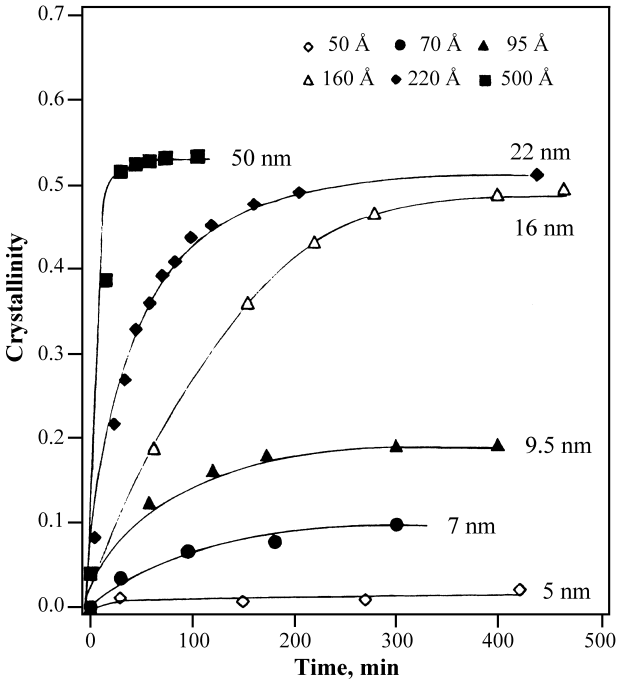


Figure 3 Kinetics of crystallization of poly(di-*n*-hexylsilane) films at 5°C for six film thicknesses between 5 and 50 nm. (Reprinted with permission from Ref. 17. Copyright 1996, American Chemical Society.)

after about 50 min, whereas for a 9.5-nm film, it takes about 250 min. In addition, the final degree of crystallinity decreases with the film thickness, with values of about 53% in the first case and 16% in the second one.

Additional results [15] indicated more precisely that the rate of crystallization starts to decrease for films thinner than 30 nm. Figure 4 illustrates the same phenomenon at a crystallization temperature of −5°C. The decrease of the kinetics of crystallization is obvious as the film thickness decreases, but, in comparison with crystallizations at 5°C, the overall degree of crystallinity and the crystallization rate increase. For example, a 7-nm film reaches a final degree of crystallinity above 50% when crystallized at −5°C as compared to 10% when crystallized at 5°C. This phenomenon can be interpreted using the concept of critical nucleus: As the crystallization temperature decreases, the supercooling increases and the size of thermodynamically stable critical nuclei decreases, allowing crystallization in thinner films.

This kinetic study demonstrates that a hydrophobic OTS-treated quartz surface enables a faster crystallization than an untreated glass substrate, but

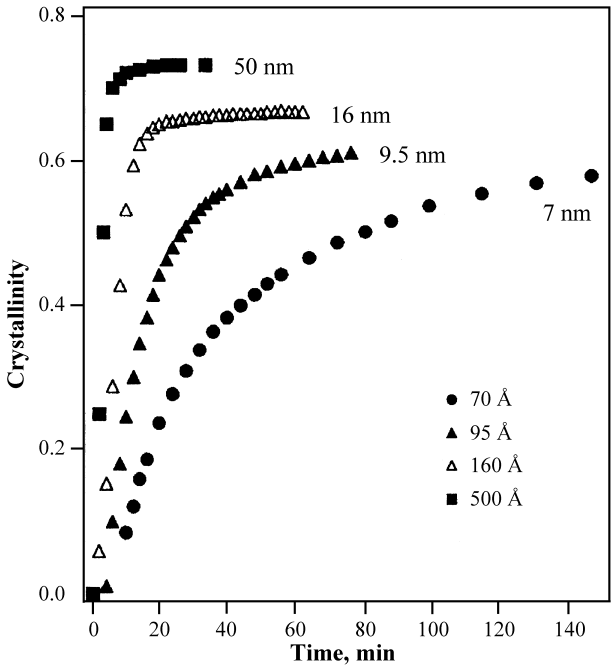


Figure 4 Kinetics of crystallization of poly(di-*n*-hexylsilane) films at short times at −5°C for four film thicknesses between 7 and 50 nm. (Reprinted with permission from Ref. 17. Copyright 1996, American Chemical Society.)

hydrophobic or hydrophilic substrates have no influence on the overall final crystallinity. Intermolecular interactions between the hexyl side chains and the substrate might be responsible for an acceleration of the secondary nucleation and subsequent growth rate of the crystallites.

A similar decrease of the degree of crystallinity with film thickness was reported in ferroelectric poly(vinylidiene fluoride-*co*-trifluoroethylene) copolymer thin films [18–20], which are considered for applications, like data storage, that require polymer films as thin as possible. Indeed, the ferroelectric response is determined by the crystallinity of the film, a low value leading to a useless film. Two different techniques were used to monitor the decrease of crystallinity. Urayama et al. prepared thin films spin-coated on aluminum foils that were selectively dissolved in a sodium hydroxide solution, leaving the undamaged copolymer film for differential scanning calorimetry (DSC) measurements [18]. Independently, Zhang et al. [19] obtained similar results with X-ray measurements: The crystallinity was found to decrease abruptly in

films thinner than a threshold thickness close to 100 nm. This behavior was correlated with a decrease of both the remnant polarization and the dielectric constant, as shown in Fig. 5. The threshold of 100 nm in thickness cannot be dictated by the radius of gyration of the polymer, which was about 15 nm, or by the lamellar thickness of the crystallites, which was about 20 nm. In addition, longer annealing times did not improve the crystallinity of the ultrathin films and the resulting ferroelectric properties.

Additional experiments by the same group [20] demonstrated that the threshold thickness (d_{th}), below which the crystallinity drops, decreases with

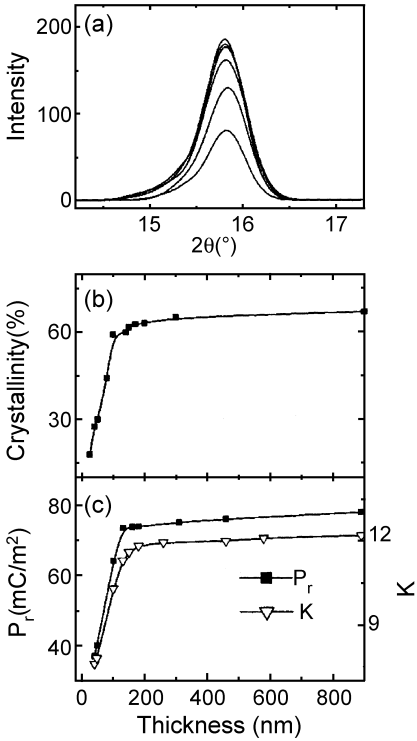


Figure 5 (a) X-ray diffraction of poly(vinylidiene fluoride-*co*-trifluoroethylene) films for six film thicknesses. From top to bottom, the thickness is 480, 200, 150, 100, 75, and 48 nm; (b) degree of crystallinity measured from X-ray diffraction as a function of film thickness; (c) remnant polarization P_r measured under a field of 160 MV/m and dielectric constant K as a function of film thickness. In all cases, the films were annealed at 140°C for 2 h. (Reprinted with permission from Ref. 20. Copyright 2002, American Institute of Physics.)

the annealing temperature, as shown in Fig. 6, which means that by lowering the annealing temperature, thinner films with good functional properties can still be obtained. Complementary X-ray diffraction measurements indicated that the polymer chains in the crystals are lying parallel to the substrate and that the (200) and (110) size of the crystals (D_{200} and D_{110}) normal to the substrate depends on the thickness of the film. From a "stable" value in thick films D_{200} and D_{110} decrease to values close to the film thickness in ultrathin films. Moreover, the "stable" crystal thickness changes with annealing temperature and always remains close to the threshold thickness, as shown in Fig. 6. This observation suggests that the threshold thickness might be determined by the size of the "stable" crystal lamellae normal to the substrate, which would mean that the crystallinity starts decreasing when the film is thinner than this "stable" crystal dimension. In the above-described system, the threshold thickness is much larger than the "classical" lamellar thickness, which is about 10–20 nm in most crystalline polymers, but is close to D_{200} and D_{110}.

In comparison, an ellipsometric study [21] showed that the melting temperature of ultrathin poly(ethylene-*co*-vinyl acetate) films decreases for films thinner than 30 nm. This phenomenon has been interpreted by the decrease of the lamella thickness when the film thickness reaches values which are of the same order of magnitude as the lamella thickness. This finding correlated with the 30-nm limit found by Frank et al. [15] at which the crystallization rate starts to decrease. However, Schönherr and Frank [22] also reported results where the lamella thickness and the melting point of poly(ethylene oxide) (PEO) do not show a significant dependence on film thickness down to 18 nm.

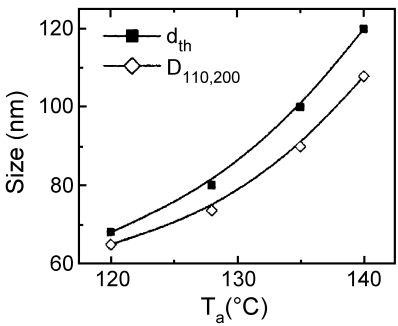


Figure 6 Threshold thickness d_{th} (observed in thin films) and "stable" crystal thickness $D_{110,\ 200}$ (observed in thick films) of poly(vinylidiene fluoride-*co*-trifluoroethylene) as a function of the annealing temperature T_a. (Reprinted with permission from Ref. 20. Copyright 2002, American Institute of Physics.)

Finally, direct observations by optical microscopy indicated a decrease of the spherulitic growth rate of PEO with a decrease of film thickness [23]. Figure 7 shows an obvious decrease when the film thickness falls below 200 nm. More precisely, the growth rates decrease by a factor of 3 for film thicknesses going from 200 to 40 nm, at crystallization temperatures ranging from 42°C to 56°C. A simple kinetic analysis indicated that the transport of the polymer chains to the growing crystals can be affected by the thickness below 200 nm, whereas the nucleation at the growth front seemed to be thickness independent. Recent results by Schönherr and Frank [22,24] confirmed those observations. Fourier transform infrared (FTIR) spectroscopy and hot-stage AFM experiments indicated a decrease of the degree of crystallinity and of the lamellar growth rates of PEO films thinner than 250 nm; at the same time, the crystallization of PEO in thin and ultrathin films was shown to follow the same laws as for bulk crystallization [i.e., the Hoffman–Weeks extrapolation procedure, the Gibbs–Thomson equation, and the Lauritzen–Hoffman theory [3,11],] and the same equilibrium melting temperatures and surface free energies of the fold surfaces. The decrease in growth rates was explained by an increase of the glass transition temperature (T_g) of the thinner films due to interactions with the interactive SiO_2 substrate.

The above-reported examples indicate that the crystallization of polymers in confined environments (i.e., in thin or ultrathin films) can decrease both the kinetics of crystallization and the crystallinity of semicrystalline polymers. The mechanisms involved are not fully understood, but the difficulties

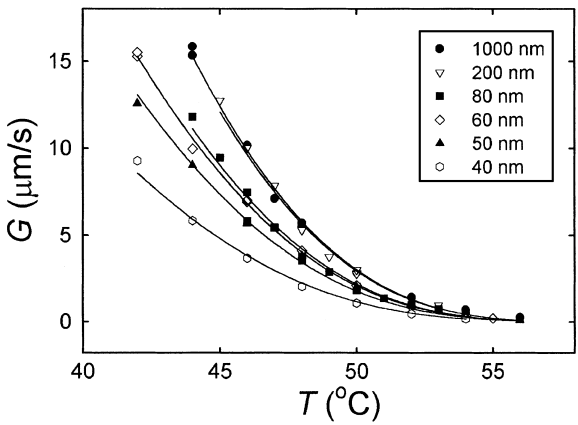


Figure 7 Radial growth rate of PEO thin and ultrathin films as a function of isothermal crystallization temperature for six film thicknesses between 40 and 1000 nm. (Reprinted with permission from Ref. 23. Copyright 2001, John Wiley & Sons, Inc.)

for polymer chains to fold in thermodynamically stable nucleus or lamellae of thicknesses close to the film thickness, associated with a possible slow-down of the diffusion of polymer chains in the melt of thin films, seem to be key factors.

III. MORPHOLOGY OF POLYMER THIN AND ULTRATHIN FILMS

In addition to growth rates and crystallinity, the orientation of crystallized polymer chains also depends on the film thickness in thin and ultrathin films, as mentioned in the previous section [14,16,20]. There seems to be a general tendency for the polymer chains to lay either exclusively parallel or perpendicular to the substrate when the film thickness is reduced, depending on the polymer–substrate interactions. Preferential edge-on lamellar orientation (polymer chains parallel to the substrate), rarely accompanied with a branching or twisting of the lamellae, was reported in two additional studies in which high-density polyethylene (HDPE) thin films, spin-coated at elevated temperatures, were observed by AFM [25,26]. In both studies, a morphological threshold thickness was found, below which flattened spherulitic structures are replaced by sheaflike aggregates of edge-on lamellae, as illustrated by Fig. 8. For Bartczak et al. [26], the threshold thickness between the two

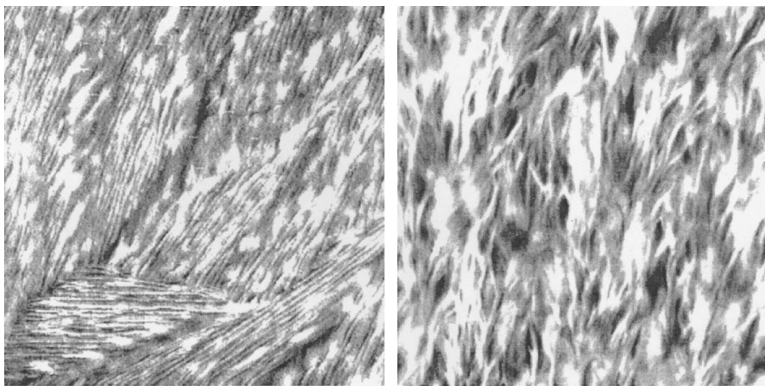


Figure 8 AFM images ($10 \times 10\ \mu m^2$) of a 50-nm (left) and 220-nm (right)-thick HDPE film spin-coated on a silicon wafer. Sheaflike structures that prevailed below a thickness of 100 nm are observed on the left image, whereas the right image shows more disordered lamellae than those present in thinner films. (Reprinted with permission from Ref. 25. Copyright 2001, American Chemical Society.)

morphologies was found to be about 300 nm, somewhat higher than the 100 nm reported by Melbring et al. [25]. Such differences are likely due to different substrate and crystallization conditions. X-ray diffraction experiments [26] suggested that the polymer-chain axis becomes parallel to the film surface (or somewhat tilted, which is consistent with the tilt angle known to exist between the chain axis and the lamella normal in polyethylene) as the film thickness decreases, confirming the edge-on orientation of the lamellae observed by AFM. In order to explain this phenomenon, a preferential formation of secondary nuclei in contact with the substrate (at the growth front of the lamellae) was proposed [26]. This could explain why, in thick films (1200 nm), some preferred crystallite orientation is still present, certainly due to the crystallites in contact with the substrate.

The presence of preferred edge-on orientations of the lamellae in thin films could be responsible for the toughening effect observed in filled HDPE when crystallized in contact with rubber and $CaCO_3$ particles [27,28]. At the surface of the particles, the polymer is likely confined, just like it is in thin films, and a specific orientation of the crystallite is generated, which is responsible for the toughening of the composite.

If edge-on lamellar morphology seems to be frequent in thin and ultrathin polymer films, flat-on lamellae have been observed in single crystals of ultrathin films. Schönherr and Frank observed that, in PEO films, the transition between edge-on to flat-on preferential orientation of lamellar crystals depends on film thickness [24], with a majority of edge-on crystals observed in films thicker than 1 μm and only flat-on crystals in films thinner than 300 nm. This phenomenon is illustrated in Fig. 9. The same group observed exclusively edge-on orientation in isotactic polypropylene (iPP) ultrathin films [29].

Single-crystal-like structures, called terraces, were found to grow in poly(ethylene terephthalate) (PET) ultrathin films [30] when the spherulitic growth stops, leaving a thin melted layer 2–4 nm thick between spherulites. At a growth rate six times lower than that of the spherulites, the 13-nm-thick terrace morphology develops in a periodic pattern characteristic of a diffusion-controlled process. An example of this so-called finger pattern is shown in Fig. 10, where direct observations by laser scanning confocal microscopy allow one to see the tip-splitting morphology that develops at the final growth stage of the terraces. According to Sakai et al., the terrace morphology appears in the ultrathin PET melt layer because, in this layer, amorphous polymer chains are disentangled and can only diffuse to the terrace growth front from one direction. This enables a very regular crystallization as compared to the disorder found during the spherulitic crystallization (observed when the melted film is thicker), where amorphous chains coming from various directions can be incorporated at the same time at the growth front.

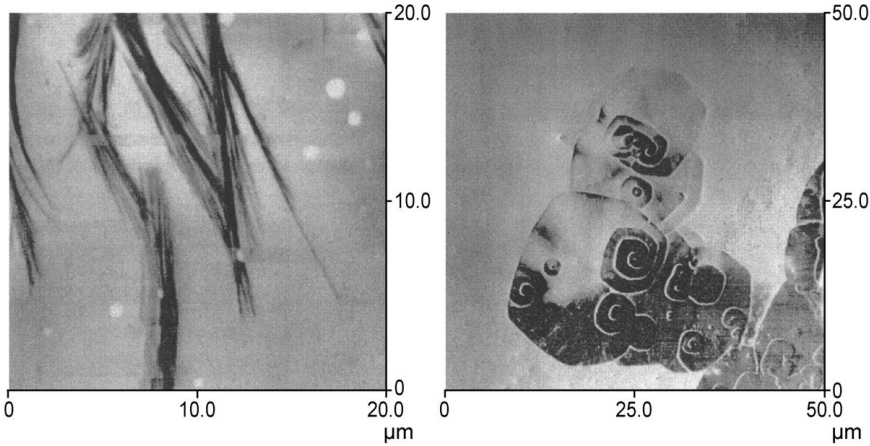


Figure 9 AFM phase images of pyrene end-labeled PEO films on oxidized silicon with thicknesses of ~2.5 μm (left) and 110 nm (right). Both images were acquired *in situ* at ~57°C, and the featureless areas correspond to the polymer melt. Edge-on lamellar crystals are observed in the thicker film (left), whereas flat-on lamellae are observed in the thinner film (right). (Reprinted with permission from Ref. 24. Copyright 2003, American Chemical Society.)

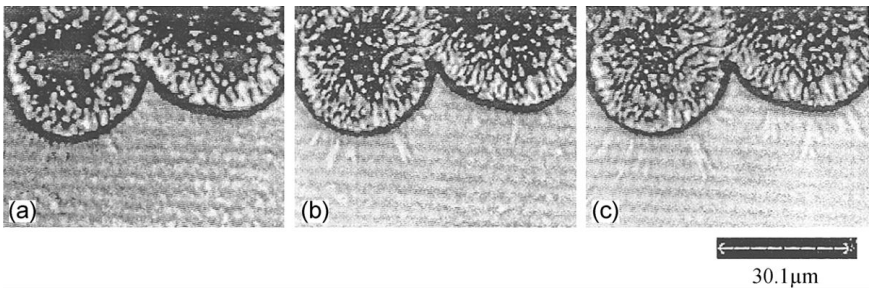


Figure 10 Time evolution of growing terraces observed by *in situ* laser scanning confocal microscopy, in PET thin films with a thickness of ~0.2 μm, isothermally crystallized at 220°C. Terraces with a fingerlike shape are observed at the periphery of the spherulites, and some of them exhibit tip-splitting. Crystallization times are (a) 1710 s, (b) 2390 s, and (c) 4420 s. (Reprinted with permission from Ref. 30. Copyright 1999, Elsevier.)

In contrast to single-crystal-like structures, true single crystals have been observed for more than 15 years in isotactic polystyrene (iPS) bulk samples crystallized at high temperatures [31,32]. IPS ultrathin films can then be used as a tool to learn more about its morphology [33–38]. Indeed, the iPS crystal morphology was not affected by the film thickness in films thicker than 20 nm [33,37], even if the crystal growth rate decreased in films thinner than 150 nm [36], with 30 nm-thick films having a growth rate about 75% of that in the bulk. In contrast, single crystals observed in iPS thin films with thicknesses above 20 nm, isothermally crystallized at 210°C, displayed a hexagonal habit with smooth, flat, lateral crystal facets similar to those observed in the bulk [32]. In ultrathin films, the morphology was shown to be temperature dependent, as it is in the bulk. As the crystallization temperature decreased, the population of overgrowth on the single crystals increased to the point of giving a densely branched morphology when isothermally crystallized at 150°C, with classical spherulites in the bulk and two-dimensional spherulites in ultrathin films [34].

Even if the iPS crystal morphology observed in 50-nm-thick films was not very different from that observed in the bulk, it is still affected by the substrate at such low thicknesses [35]. As illustrated in Fig. 11, crystals developed around giant screw dislocations when the ultrathin films were supported by mica, glass, or Kapton, and around small bundles of edge-on lamellae (which grew perpendicular to the plane of the film) when growing on carbon or silicon. The two morphologies were explained by two different types of primary nucleation, acting either on the substrate surface (giving screw dislocations) or on the "core" of the melt (giving bundles of lamellae), depending on the interactions between the substrate and the melt. However, as the density of screw dislocations (density of overgrowth on already crystallized mother crystal) was higher onto the Kapton substrate, it was suggested that not only the primary nucleation is influenced by the substrate but that most of the molecules "feel" the influence of the substrate due to the low thickness of the films.

Recent findings by the same group [37] indicate that different morphologies are found when decreasing the film thickness down to 11 nm. Figures 12 and 13 illustrate this morphological evolution, with AFM images of crystals found in films thicker than 50 nm shown in Fig. 12 and the new morphologies observed (also by AFM) in 11-nm-thick films in Fig. 13. Whereas films thicker than 50 nm exhibit many overgrowth lamellae, the crystals grown in 11-nm-thick films consist of one single lamella, as demonstrated by the single crystallographic orientation found by electron diffraction.

Moreover, the morphological changes observed in Fig. 13 when decreasing the crystallization temperature in 11-nm ultrathin films do not correspond to the phenomena observed in Fig. 12 in thicker films. The branched morphol-

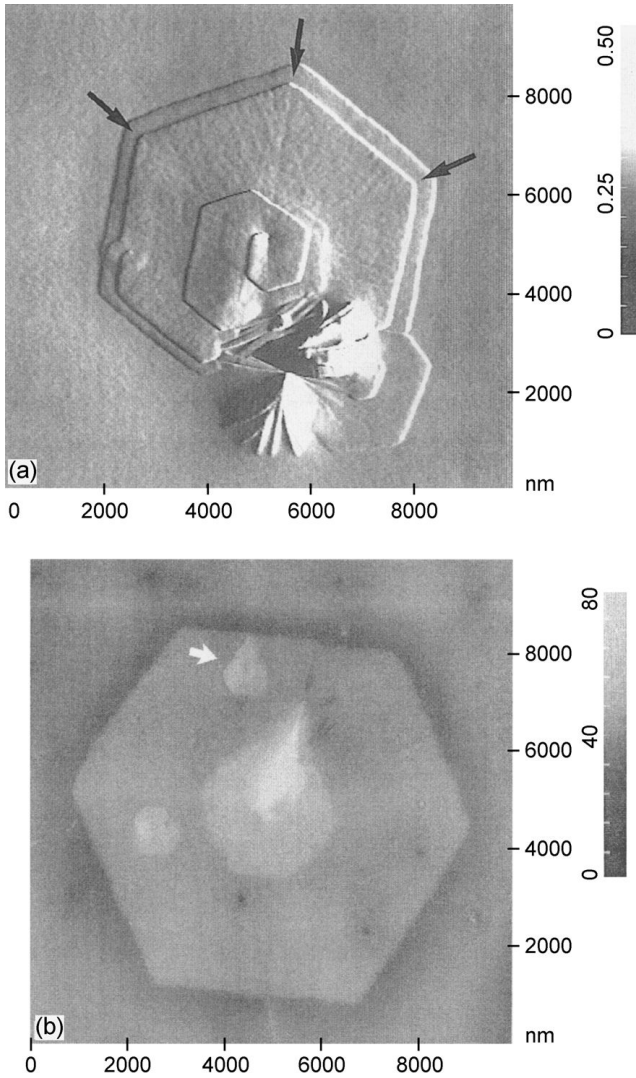


Figure 11 (a) Contact-mode AFM (force mode) image of an isotactic PS crystal grown from the melt in a 50-nm-thick film at 210°C on a mica substrate. The arrows indicate sector boundaries. The crystal display a giant screw dislocation at the center. (b) Height-mode image of a lamella crystallized under similar conditions on a carbon substrate. The arrow indicates an elongated overgrowth terrace. The crystal has developed around a bundle of lamellae. (Reprinted with permission from Ref. 35. Copyright 1997. Kluwer.)

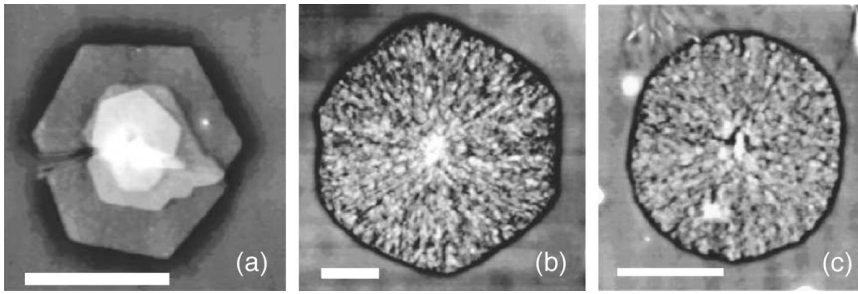


Figure 12 AFM images of isotactic PS crystals grown in films thicker than 50 nm at (a) 210°C for 2 h 45 min (hexagonal plates), (b) at 195°C for 1 h (rounded hexagonal plates), and (c) at 180°C for 20 min (disks). Scale bars represent 5 μm. Each of these crystals are made of many overgrown lamellae. (Reprinted with permission from Ref. 37. Copyright 2001, Elsevier.)

ogies observed in the 11-nm ultrathin films correspond to a diffusion-controlled phenomenon giving dendrites and compact seaweed structures, whereas the hexagonal morphology observed at higher crystallization temperatures corresponds to a classical nucleation-controlled phenomenon. As the crystallization temperature decreases, the lateral growth interface of the original hexagonal morphology becomes unstable, this instability resulting from the gradient of film thickness around the crystal, as observed by AFM. This thickness gradient, always observed by AFM around polymer crystals [39], has a huge morphological influence in the present system due to the very thin films that were observed. The gradient is created because polymer crystals develop to become thicker than the original film and, then, protrude from the film surface, leaving a depletion area around the growing crystal [35].

Additional experiments [38] indicated that the crystal growth rate of iPS decreases drastically in films thinner than the lamellar crystal. For example, 8-nm-thick lamellar crystals developing in 5-nm-thick films have a growth rate corresponding to 5% only of the bulk growth rate. In such films, the dendritic morphology of the crystals becomes more open as the film thickness decreases and the mean width of the branches of the dendrites becomes larger. Poly[(*S*)-lactide] single crystals, isothermally crystallized at 160°C in ultrathin films, exhibit the same transition from hexagonal to dendritic when the film thickness is decreased [40]. However, the transition takes place at higher thicknesses, about 40 nm, and no growth rate difference was observed between the two morphologies.

Diffusion of the polymer chains seems to control the morphology of ultrathin polymer films, in which fingerlike and fractal patterns are often observed, (e.g., during the crystallization of low-molecular-weight PEO

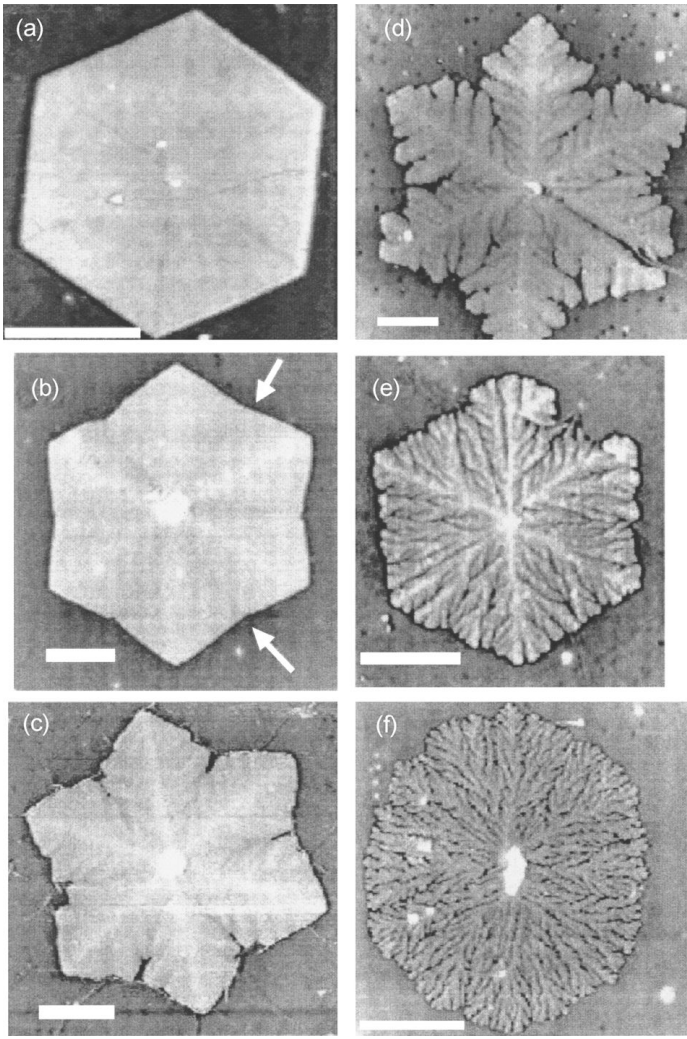


Figure 13 AFM images of isotactic polystyrene crystals grown in 11-nm-thick films (a) at 210°C for 4 h, (b) at 205°C for 4 h, (c) at 200°C for 4 h, (d) at 195°C for 3 h, (e) at 190°C for 1 h, and (f) at 180°C for 1 h. Scale bars represent 5 μm. The arrows in (b) show the reentrant sites, where the growth rates decrease during growth. (Reprinted with permission from Ref. 37. Copyright 2001, Elsevier.)

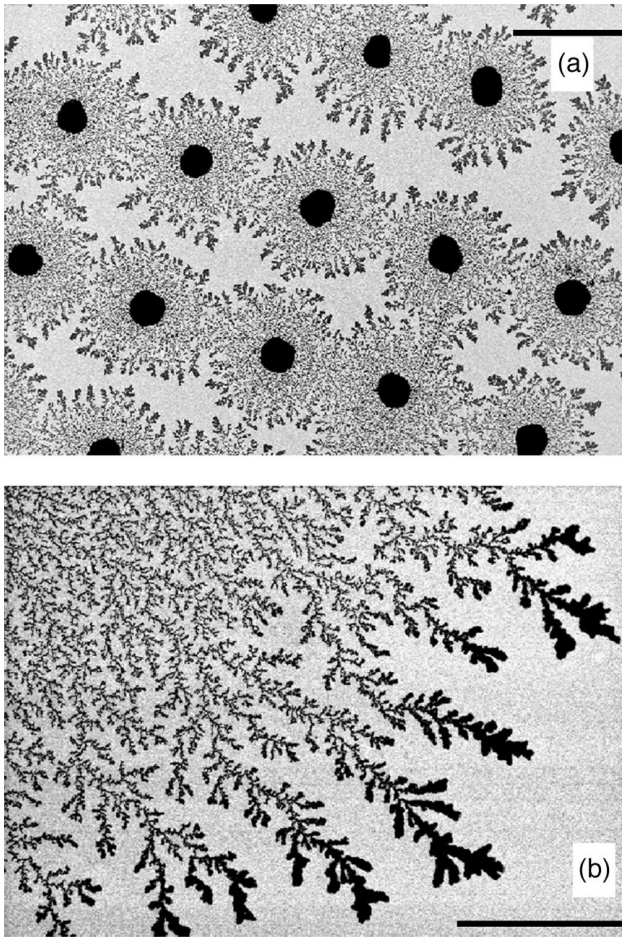


Figure 14 Scanning electron micrographic images: (a) Branched crystalline patterns emanating from PEO (M_w = 6000 g/mol) dots on Au surfaces; (b) magnified view in the region near the pattern periphery of a PEO (M_w = 2000 g/mol) dot. Samples were crystallized at 22–24°C for 24 h in 53% relative humidity. The scaling bars are equal to 20 μm in (a) and 10 μm in (b). (Reprinted with permission from Ref. 42. Copyright 2002, Wiley–VCH.)

[41,42]). Reiter and Sommer [41] observed this diffusion-controlled morphology in 5-nm-thick PEO monolayers resulting from a pseudodewetting process, whereas Wang et al. [42] made similar observations in ultrathin PEO films created by chains diffusion (in condensed water) from thick PEO dots. Finally, Schönherr and Frank observed the same phenomenon in higher-molecular-weight PEO crystallized in films thinner than 15 nm [24].

In both studies [41,42], primary nucleation appears in the thicker part of the film leading to the dendritic crystallization of the adsorbed PEO monolayers. Reiter and Sommer [41] found that the growth rate of the fingerlike patterns is about 10 times slower than the growth rates measured in the bulk. Wang et al. [42] observed a fractallike morphology [43], close to the PEO dots, then a dense branchinglike morphology [44] when the supply of PEO chains become limited, away from the PEO dots, as shown in Fig. 14. The PEO lamellae were found to be flat-on in both studies, with the polymer chains perpendicular to the substrate, as for the dendrites observed in ultrathin iPS films. A depletion zone just ahead of the growing crystals was considered by Reiter and Sommer responsible of the particular morphologies in very thin films. At low crystallization temperatures, as the growth rate increases, the depletion zone becomes smaller and the crystal front unstable, leading to fingerlike patterns of small widths, whereas the opposite is observed when the crystallization temperature is close to the melting point. These observations are similar to those made on iPS dendrites, whose branch width becomes larger when the film thickness is decreased (along with the crystal growth rate) [38].

All of these studies indicate that the film thickness has a large influence both on the orientation and on the final shape of the lamellae. From spherulites in which lamellae have a random orientation (relative to the substrate) in thick and thin films, the crystalline organization evolves toward single crystals with film thicknesses down to ultrathin dimensions, the lamellae being oriented either flat-on or edge-on, depending on the polymer–substrate interactions. Upon a further decrease of the film thickness, below a threshold value close to the thermodynamically stable lamellar thickness, dendrites can be observed as the lamellar growth front becomes unstable, probably due to the presence of a thickness gradient around the growing crystals.

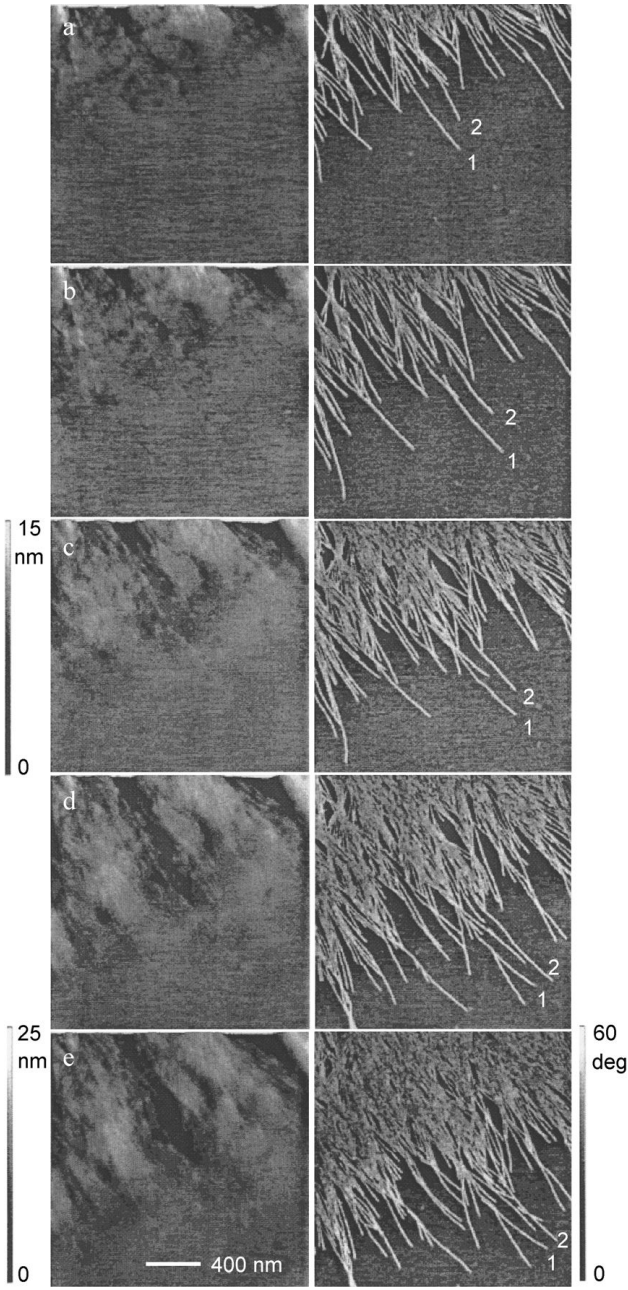
IV. THIN FILMS AS A TOOL FOR THE UNDERSTANDING OF THE CRYSTALLIZATION OF POLYMERS

The first AFM observations of the growth of spherulites in a thin polymer film were reported in 1998 by Hobbs et al. [39]. In that study and others [45–47], thin and ultrathin samples were used to reach a better understanding of the

polymer crystallization phenomenon. Indeed, the poly(hydroxybutyrate-*co*-hydroxyvalerate) spherulites observed by AFM during crystallization at room temperature revealed, for the first time, that the lamellae at the growth front, called dominant lamellae, do not grow at a constant rate [39]. Even if, at a larger scale, the growth rate of spherulites is constant, dominant lamellae were found to grow faster than the spherulites, before slowing down and even stopping. These observations, similar to recent findings by the same group on shish kebab crystallization of polyethylene in ultrathin films [45], are in contradiction with classical theories of crystallization [11]. Hobbs et al. [39] then proposed that the growth rate of individual lamellae does not control the growth rate of the spherulites, which is, instead, controlled by the nucleation rate of new lamellae on already dead ones. These conclusions were recently challenged by Schönherr and Frank [22], who observed a constant linear growth rate of PEO single lamellae growing in ultrathin films.

In the above-mentioned studies, the film thickness is an important parameter not only because the reduced thickness leads to a different morphology from that observed in the bulk but also because thin films allow a better visualization of the crystalline morphology (by AFM, for example). In thin films, all of the lamellae grow in the same plane, contrary to thick films, where the lamellae found at the surface may originate from a primary nucleus positioned everywhere, even at the bottom of the film. In addition, the roughness of the film surface is usually attenuated in thin films, which makes AFM experiments easier. Finally, as illustrated in the previous section, lamellae tend to grow exclusively with edge-on or flat-on orientation when the film thickness is decreased.

For example, most of the lamellae of poly(bisphenol A octane ether) observed by AFM phase imaging in 300-nm thin films were oriented edge-on [46,47]. This particular orientation proved to be very useful for measuring the growth rate of individual lamellae, their nucleation, and organization. Direct observations of the formation of primary nuclei were reported [46], in which the instability of small primary nuclei, as predicted by thermodynamics, was detected by AFM phase imaging, and nuclei smaller than the critical dimensions were seen to vanish instead of giving birth to another lamella. In agreement with the observations made by Hobbs et al., lamellae were not found to grow at a constant rate, as shown in Fig. 15, where lamellae 1 and 2 can be followed during their growth. For the first time, induced nuclei giving birth to daughter lamellae (responsible for the branching) were observed. They usually develop between 0.5 and 1 μm behind the tip of the growing lamellae. Li et al. [47] assumed that when a loose loop or protruding cilia is trapped in the lattice of the parent lamellae, the mobility of the polymer chain is decreased, which leads to the formation of an induced nucleus. Some daughter lamellae were even found to grow in the direction opposite to the growth of the spherulite, but they usually grow at a slower rate than the



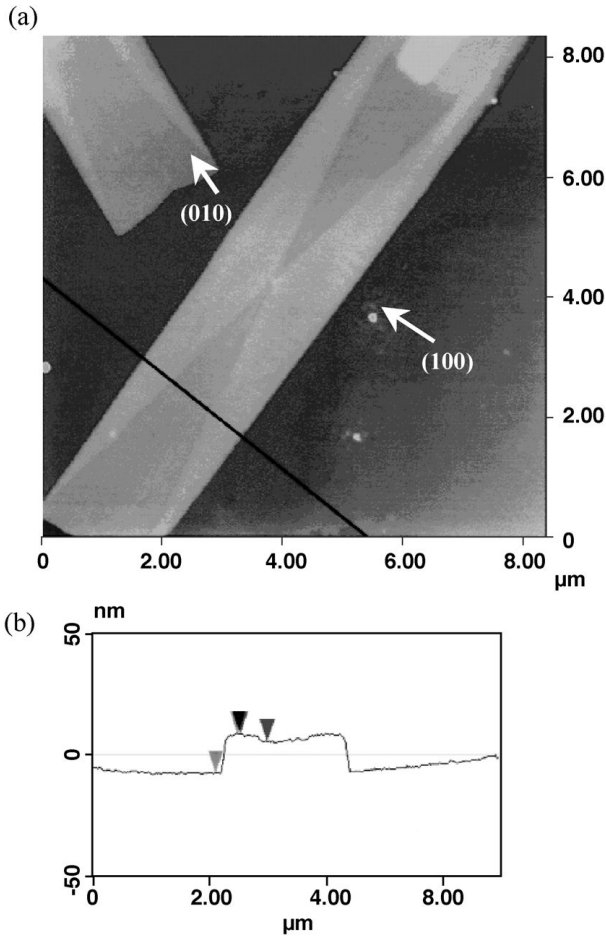


Figure 16 (a) Tapping-mode AFM height image of a syndiotactic polypropylene lamellar single crystal isothermally grown from the melt at 130°C in a film 25–50 nm thick. (b) Morphological profile corresponding to the black line drawn in the image. Thicknesses measured on this profile are 20.6 nm in the (100) sector and 17.5 nm in the (010) one. (Reprinted with permission from Ref. 48. Copyright 2000, American Chemical Society.)

Figure 15 Sequence of AFM images of the lamellae at the growth front of a spherulite in a poly(bisphenol A octane ether) polymer thin film (300 nm), crystallized at 25°C. At each of the five consecutive times (time interval ~9 min), both height (left) and phase (right) images were recorded simultaneously. Data scale of the height images: 15 nm (a–c) and 25 nm (d, e). (Reprinted with permission from Ref. 47. Copyright 2001, American Chemical Society.)

lamellae growing in the direction of the surrounding melt, certainly due to the lack of free polymer chains. In general, the growing tip of the lamellae were found to have a lower phase shift than the body of the lamellae, close to what is observed for the melt in phase images. Thus, the growth front appears softer than the body of the lamella and can be interpreted as a kind of transition between amorphous and crystalline states.

Other polymers, like iPS and syndiotactic polypropylene (sPP), can lead to flat-on isolated lamellae called single crystals in ultrathin films having thicknesses of about 50 nm when crystallized at low supercoolings [33,48]. These single crystals, grown from the melt, can be used as models to study the crystallization of polymers. For example, the lamellar thickness of iPS single crystals was measured accurately using AFM height images [33] after permanganic etching to remove selectively the amorphous layer that usually covers single crystals and prevents direct measurement of lamellar thickness. After etching, the outlines of the hexagonal crystals appeared sharper in AFM height images and allowed an accurate evaluation of the lamellar thickness when planar crystals with at least one overgrowth were chosen (in order to measure the step height between two lamellar crystals).

In another example, the growth of sPP lamellar single crystals in ultrathin films was directly observed using an AFM coupled with a hot stage [48]. As shown in Fig. 16, the single crystals have a rectangular shape and are composed of four sectors: two (100) sectors thicker than the two (010) sectors. The use of ultrathin films instead of solutions to grow the single crystals enabled a direct determination of the crystal growth rate and of the thickness variation of the different sectors during crystallization. Melting observations were also possible in real time, showing that the (100) thicker sectors are the first ones to melt, but can be recrystallized when held at the melting temperature.

In summary, thin and ultrathin films can be valuable tools for studying polymer crystallization because they display a simpler morphology than bulk crystallized samples. They recently allowed one to measure growth rates at the lamellar scale in both edge-on and flat-on orientations, to observe the birth of primary and induced nuclei (resulting in branching), and to measure lamellar thicknesses. These examples illustrate how promising this field of research can be to achieve a better understanding of the polymer crystallization.

V. CRYSTALLIZATION OF MISCIBLE PCL/PVC BLENDS

Semicrystalline polymer blends are often composed of a semicrystalline and an amorphous polymer. The miscibility [49], morphology [50], and kinetics of crystallization [51] of the poly(ε-caprolactone)/poly(vinyl chloride) (PCL/

PVC) blend, which can be considered as a reference system, are well established for bulk samples, generally crystallized at medium or high supercoolings. In the melt, the blend is fully miscible, and when PCL lamellae develop in the crystalline phase, the miscibility is still observed in the amorphous phase. In a recent study, our group investigated the influence of the film thickness on the growth rate of PCL/PVC blends [52], and observed two types of spherulites in thick PCL/PVC films crystallized at low supercoolings [52,53].

As shown in Fig. 17, the growth rate of PCL/PVC thin films of 80/20 composition decreases with the film thickness between 37°C and 49°C, but this seems to be a common feature for polymer films [15,22–24,30,36,38,41]. However, in most of the examples reported in the literature, the decrease in growth rate was observed in ultrathin films, not for films with thicknesses above 200 nm. In contrast, in the PCL/PVC blend, the growth rate decreased by about 20% when the film thickness decreased from 800 to 300 nm, and by about 50% when it further decreased to 100 nm.

Similar decreases in growth rates were observed for all PCL/PVC compositions from 60/40 up to 80/20, whereas the growth rates of pure PCL films were insensitive to the film thickness for the range of thicknesses studied.

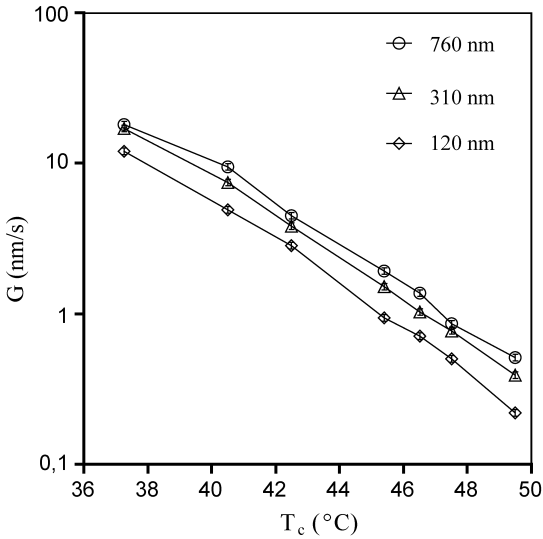


Figure 17 Radial growth rate of a 80/20 PCL/PVC blend as a function of crystallization temperature for film thicknesses of 120, 310, and 760 nm. (Reprinted with permission from Ref. 52. Copyright 2003, American Chemical Society.)

Therefore, specific properties of the polymer blends must be invoked to explain these observations.

Complementary observations of thick PCL/PVC films crystallized at low supercoolings highlighted another specific behavior of this blend. As shown in Fig. 18 (Mareau, unpublished data), two types of spherulites that nucleated and developed at different depths in the film, with different morphologies and growth rates, are observed in a PCL/PVC 30-μm-thick film of 60/40 composition. Spherulites that nucleated at the open surface of the thick film were labeled S-pherulites (S standing for surface), and they developed faster than the other type of spherulites, which nucleated in the bulk and were labeled C-spherulites (C standing for core). For the example shown in Fig. 18, the S-spherulite crystallizes more than three times faster than the C-pherulite; this difference was found to decrease for blends with higher PCL content.

Morphologies of the two types of spherulite are also very different: S-Spherulites display an open structure with large radial fibrils, a wavy growth front, and low birefringence, whereas C-pherulites are compact with concentric rings, a circular growth front, and the classical banded Maltese Cross pattern when observed between crossed polars. The two types of spherulite were observed at several blend compositions from 60/40 up to 80/20, but only for crystallization temperatures above a limiting value that was found to increase with the PCL content in the blend.

Comparable dual growth rates and morphologies were observed in homopolymer thick films [54,55]. In a study on the crystallization of a poly(ether ether sulfide), Srinivas et al. [54] described two types of spherulite which are in many ways similar to the S-spherulites and C-spherulites described earlier. However, they never observed the two types of spherulite growing in the same area of the thick films. They associated the growth of spherulites of type II (similar to S-spherulites) to a film thickness of about 10 μm and spherulites of type I (similar to C-spherulites) to a film thickness above 50 μm, without giving any reason for this difference.

Finally, poly(dimethylsiloxane) spherulites with bright or dark birefringent patterns, coexisting in the same area, were reported by Singh [55] in thick films (10–25 μm) when observed between crossed polars. The bright spherulites display a positive birefringence with a classical Maltese Cross extinction pattern, whereas the birefringence sign of the dark spherulites changes from negative in their center to positive in the outer area, with an annular transition of low birefringence, as illustrated in Fig. 19. In addition, dark spherulites retain a constant growth rate through the change of birefringence sign, lower than that observed for bright spherulites. The crystallization temperature does not affect the proportion of dark spherulites, estimated to be about 10–20% of the field of view. Dark spherulite morphology was interpreted as the growth of dominant lamellae with a negative birefringence in the center, later compensated by the branching of secondary lamellae with positive birefrin-

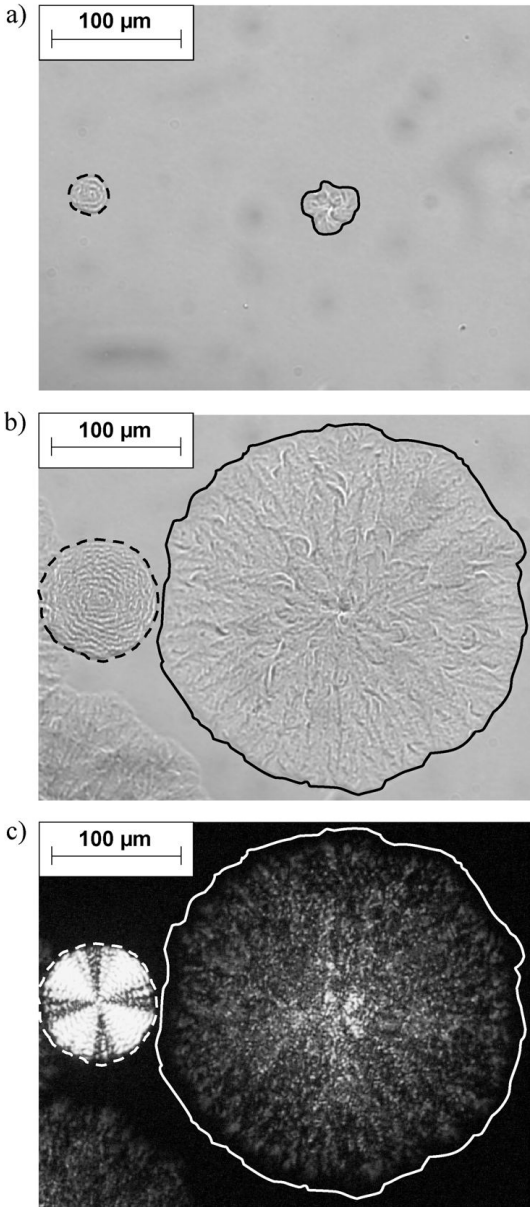


Figure 18 Optical microphotographs of S-type (circled by a solid line) and C-type (circled by a dotted line) spherulites for a 30-µm-thick film of a 60/40 PCL/PVC blend, crystallized at 40°C: (a) bright-field image taken at 30.07 h; b) bright-field image, same position, 51.15 h later; (c) crossed polar image, same position a few seconds after image b). (V. H. Mareau, unpublished data.)

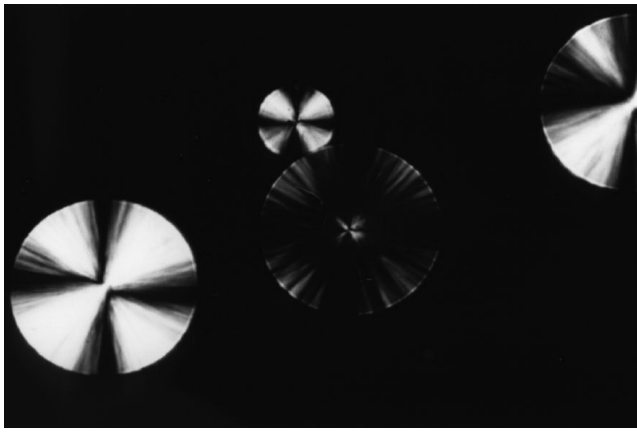


Figure 19 Crossed polar optical microphotograph of a poly(dimethylsiloxane) thick film, 10–25 μm in thickness, isothermally crystallized at −65°C. A dark spherulite with a bright center and outer regions, and a dark annulus, is surrounded by three bright spherulites. (Reprinted with permission from Ref. 55.)

gence (the zone of perfect compensation giving the dark annulus), that finally give the positive birefringence of the outer section of the dark spherulites.

X-ray photoelectron spectroscopy (XPS) experiments performed on both thin and thick PCL/PVC films indicated that the free surface of the films is composed of almost pure PCL. The presence of an enriched PCL surface, which can be associated with the lower surface tension of PCL [56], can explain, at least in part, both the variation of the spherulitic growth rate with the film thickness in thin films and the development of two different types of spherulite. Indeed, because the presence of PVC slows down the growth rate of PCL, spherulites nucleated in a PCL-enriched layer can crystallize faster than those nucleated in the bulk of the PCL/PVC films. In addition, there are similarities, at each blend composition, between the growth rates of C-spherulites and spherulites in 100-nm thin films, on the one hand, and S-spherulites and spherulites in 800-nm thin films, on the other hand. These similarities were interpreted in terms of the absence of influence of the PCL surface enrichment in the development of spherulites in 100 nm thin films in the first case and of a maximum effect of the PCL surface enrichment for films as thick as 800 nm in the second case.

In a very different system, De Cupere and Rouxhet [57] observed the selective crystallization of the free surface of 3.7-μm-thick PET films. Unusual spherulites with a sectorlike shape were found to develop with a granular aspect only at the surface of the films. In this system, which is not a polymer

blend but a homopolymer, the results were interpreted in terms of a decrease of T_g near the surface, which would make crystallization easier at the surface than in the bulk. In a previous study, Hayes et al. [58] already observed, by XPS and FTIR, a higher crystallization rate at the surface (first 5 nm) of 150-nm-thick PET films. This example illustrates how the polymer film surface can behave differently from the bulk, although no direct comparison can be made with the PCL/PVC blend thick films described before. In Refs. 57 and 58, a fractionation of the sample leading to lower-molecular-weight chains at the surface (because they exhibit a lower surface tension) could explain the observed phenomena.

These examples highlight the influence of the free film surface on the crystallization of polymer films. The surface composition can be very different from the bulk composition due to an enrichment in one of the two polymers involved, as in the PCL/PVC system, or, in a homopolymer, due to a molecular-weight fractionation, which usually leads to a higher concentration of low-molecular-weight species at the surface, as observed with PET [57]. This behavior can be compared to some of the observations reported in the previous sections, where it has been seen that the substrate on which ultrathin films stand can deeply influence the crystallization and the morphology of the polymer involved due to polymer–substrate interactions.

VI. CONFINED POLYMER CRYSTALLIZATION IN BLOCK COPOLYMER FILMS

Thin and ultrathin films are examples of systems in which the polymer crystallization is confined into restricted geometries but they are not unique. Indeed, block copolymers with semicrystalline and amorphous blocks can organize into microphase-separated structures like spheres, cylinders, gyroid, and lamellae below the order–disorder transition temperature (T_{ODT}) [59], as illustrated in Fig. 20. To avoid a disruption of the mesophase during the

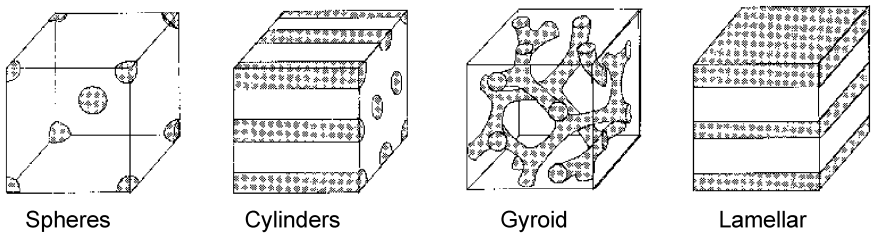


Figure 20 Schematics of block copolymer mesophase structures at equilibrium. (Reprinted with permission from Ref. 59. Copyright 1999, John Wiley & Sons, Inc.)

crystallization of the semicrystalline blocks, the copolymers need to be designed so that the T_g of the amorphous block is higher than the melting temperature (T_m) of the crystals formed by the crystallizable block, but lower than the T_{ODT} to allow the formation of the mesophase.

For example, in poly(ethylene-*b*-vinylcyclohexane) (PE–PVCH) block copolymers, the high T_g of the PVCH block limits the crystallization of the PE blocks in the confined space dictated by the mesophase structure [60]. In this study, Weimann et al. found that the confinement of the PE domains decreases the crystallite size in both the diblock and triblock copolymers, as observed by the reduction of T_m, and that the crystallite size of the doubly-ethered PE in the triblocks is smaller than that observed in the diblocks. In addition, the supercooling necessary to observe the maximum degree of crystallinity increases when moving from a gyroid sample, where there is continuity in the PE matrix, to lamellar or cylindrical samples, where PE domains are isolated. Obviously, in a continuous PE matrix, a high degree of crystallinity can be achieved with fewer nucleus than in a discontinuous one.

In another study where, again, the polymer crystallization is not constrained by the film thickness but by the mesophase morphology of block copolymers, Zhu et al. reported the crystallization of PEO blocks in a lamellar-forming poly(ethylene oxide-*b*-styrene) (PEO–PS) diblock copolymer [61,62]. In this system, the glassy PS layers were found to constrain the crystallization of the PEO into a lamellar morphology, with a thickness of about 9 nm, whereas the lamellar thickness achieved in unconstrained comparable PEO homopolymer is about 15 nm. The constrained crystallization resulted in different crystal orientations of the c axis as a function of the crystallization temperature (T_c) as observed by small-angle X-ray scattering and wide-angle X-ray diffraction. The crystals were oriented parallel to the lamellar plane, then inclined, and, finally, perpendicular as T_c increased. It was found than this orientation is not determined by the primary nucleation but fixed in the early stages of crystallization, probably when the crystal size is large enough to feel the constrained environment of the confined lamellae.

Finally, a combination of constrained geometries due to a mesophase lamellar morphology and, in addition, to the use of thin film were studied by Hong et al. in poly(ethylene oxide-*b*-1,4 butadiene) (PEO–PBD) diblock copolymers thin and ultrathin films (150–40 nm) [63]. They observed that the crystallization of the PEO layers does not destroy the lamellar morphology (parallel to the substrate) and that the PEO crystalline chains are oriented perpendicular to the lamellar layers. Moreover, the crystallization in one PEO layer was found to induce the crystallization in adjacent layers with the same crystallographic orientation. This phenomenon was explained by the presence of screw dislocations that would propagate the growth of a single crystal to adjacent layers separated by a 10-nm amorphous PBD layer.

In summary, block copolymers mesophases can be used to confine semicrystalline polymers efficiently and to study the influence of this confinement on the crystallization. As illustrated in this section, those restricted geometries affect the crystallization (e.g., by decreasing the crystallite size). Specific polymer-chain orientations resulting from constrained crystallizations were also observed.

VII. CONCLUSIONS

This review illustrates that the degree of crystallinity, the rate of crystallization, and the morphology of semicrystalline polymers are significantly modified when the crystallization takes place in a confined environment. In most studies, two-dimensional crystallization was achieved by working in thin and ultrathin films. However, block copolymers mesophases [59–63] also lead to a similar confined environment. Additional examples include HDPE droplets suspended in immiscible solvents [64], iPP droplets confined in a PS matrix [65], PEO embedded in organic networks of poly(trimethopropane trimethacrylate) [66], or polymers blended with porous inorganic materials having a layered silicate structure [67], like PEO–montmorillonite blends [68], and, finally, HDPE blended with fillers like $CaCO_3$ particles [28].

From the present overview, it clearly appears that a decrease of the film thickness, and more generally a confinement of the crystallization, leads to a decrease of the degree of crystallinity and of the rate of crystallization [16–20,22–24,36,41,52,60]. The threshold thickness below which the crystallization is hindered varies from one system to another: from a high value of 800 nm in PCL/PVC blends where the spherulitic growth rate starts to decrease [52] to a low value of 30 nm in the case of pol(di-*n*-hexylsilane) [16,17]. Between these maximum and minimum values, a threshold thickness of about 200 nm was determined for PEO [22–24], around 150 nm for iPS [36], and 100 nm for the poly(vinylidiene fluoride-*co*-trifluoroethylene) [18–20], below which, in each case, the rate of crystallization and/or the degree of crystallinity decrease. Except for poly(di-*n*-hexylsilane), the threshold thickness is always much larger than the radius of giration or the lamellar thickness of the polymer. For PEO [22,24,41] and iPS [38], growth rates were found to decrease again, but more severely this time, at film thicknesses below the bulk lamellar thickness.

In contrast to these systematic observations, no general trend was found concerning the influence of the film thickness on the orientation of the polymer chains (and of the crystalline lamellae) relative to the substrate. Specific interactions between each type of substrate and each polymer appear to determine the exclusive orientation observed in each system. Indeed, when the film thickness decreases, only edge-on [14,16,20,25–28,35,45–47] or flat-

on [30–42,48] lamellae are observed with polymer chains respectively parallel or perpendicular to the substrate. The exclusive flat-on or edge-on orientation of the lamellae in ultrathin films can then be considered as a general trend for such materials.

In studies involving PEO [24,41,42], iPS [37,38], and poly[(*S*)-lactide] [40] in ultrathin films in which flat-on lamellae is the exclusive morphology, a further decrease of the film thickness below 15, 8, and 40 nm, respectively, creates instabilities in the growth fronts and the appearence of dendritic morphologies. A depletion area around the growing crystal, due to the reduced thickness of the films and to a growth rate higher than the possible diffusion of the polymer chains to the growth front, seems to be responsible for these instabilities.

If the polymer–substrate interface has a great influence on the crystallization morphology and kinetics in ultrathin films, the observation of dual morphologies and growth rates in thick PCL/PVC films [52] due to the influence of a PCL surface enrichment added to the crystallization of PET thick films limited to the surface [57], remind us that the free film surface plays a dominant role.

Finally, the importance of thin and ultrathin films in our fundamental understanding of crystallization mechanisms was illustrated [22,33,39,45–48]. The development of new confined systems like block copolymer mesophases, associated with improvements of the available tools to probe the crystallization process in real time, like AFM, should lead to a better understanding of the crystallization process and, possibly, to significant modifications of the existing theories.

ACKNOWLEDGMENTS

This contribution was supported by grants from the Formation de Chercheurs et l'Aide à la Recherche (FCAR) Fund (Québec) and the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

1. Keller, A. Phil. Mag. 1957, *2*, 1171–1175.
2. Vaughan, A.S.; Bassett, D.C. In *Comprehensive Polymer Science: The Synthesis, Characterization, Reactions & Applications of Polymers*; Booth, C., Price, C., Eds.; Pergamon Press: Oxford, 1989; Vol. 2, 415–457.
3. Bassett, D.C. *Principles of Polymer Morphology*; Cambridge University Press: New York, 1981.
4. Woodward, A.E. *Atlas of Polymer Morphology*; Hanser Publisher: New York, 1988.

5. Keith, H.D.; Padden, F.J., Jr. J. Polym. Sci. 1959, *39*, 101–122.
6. Keller, A.; Sawada, S. Makromol. Chem. 1964, *74*, 190–221.
7. Norton, D.R.; Keller, A. Polymer. 1985, *26*, 704–716.
8. Li, J.X.; Cheung, W.L. J. Appl. Polym. Sci. 1999, *72*, 1529–1538.
9. Abo el Maaty, M.I.; Hosier, I.L.; Bassett, D.C. Macromolecules 1998, *31*, 153–157.
10. Armitstead, K.; Goldbeck-Wood, G. Adv. Polym. Sci. 1992, *100*, 219–312.
11. Hoffman, J.D.; Davis, G.T.; Lauritzen, J.I., Jr. *Treatise on Solid State Chemistry*; Hannay, N.B., Ed.; Plenum: New York, 1976; Vol 3, 497–614.
12. Hoffman, J.D.; Lauritzen, J.I., Jr. J. Res. Natl. Bur. Std. (US) 1961, *65A*, 297–336.
13. Godovsky, Y.K.; Magonov, S.N. Langmuir 2000, *16*, 3549–3552.
14. Prest, W.M., Jr.; Luca, D.J. J. Appl. Phys. 1980, *51*, 5170–5174.
15. Frank, C.W.; Rao, V.; Despotopoulou, M.M.; Pease, R.F.W.; Hinsberg, W.D.; Miller, R.D.; Rabolt, J.F. Science 1996, *273*, 912–915.
16. Despotopoulou, M.M.; Miller, R.D.; Rabolt, J.F.; Frank, C.W. J. Polym. Sci. Polym. Phys. Ed. 1996, *34*, 2335–2349.
17. Despotopoulou, M.M.; Frank, C.W.; Miller, R.D.; Rabolt, J.F. Macromolecules 1996, *29*, 5797–5804.
18. Urayama, K.; Tsuji, M.; Neher, D. Macromolecules 2000, *33*, 8269–8279.
19. Zhang, Q.M.; Xu, H.; Fang, F.; Cheng, Z.-Y.; Xia, F. J. Appl. Phys. 2001, *89*, 2613–2616.
20. Xia, F.; Razavi, B.; Xu, H.; Cheng, Z.-Y.; Zhang, Q.M. J. Appl. Phys. 2002, *92*, 3111–3115.
21. Kim, J.H.; Jang, J.; Zin, W.-C. Macromol. Rapid Commun. 2001, *22*, 386–389.
22. Schönherr, H.; Frank, C.W. Macromolecules 2003, *36*, 1199–1208.
23. Dalnoki-Veress, K.; Forrest, J.A.; Massa, M.V.; Pratt, A.; Williams, A. J. Polym. Sci. Polym. Phys. 2001, *39*, 2615–2621.
24. Schönherr, H.; Frank, C.W. Macromolecules 2003, *36*, 1188–1198.
25. Mellbring, O.; Kihlman Øiseth, S.; Krozer, A; Lausmaa, J.; Hjertberg, T. Macromolecules 2001, *34*, 7496–7503.
26. Bartczak, Z.; Argon, A.S.; Cohen, R.E.; Kowalewski, T. Polymer. 1999, *40*, 2367–2380.
27. Bartczak, Z.; Argon, A.S.; Cohen, R.E.; Kowalewski, T. Polymer. 1999, *40*, 2331–2346.
28. Bartczak, Z.; Argon, A.S.; Cohen, R.E.; Kowalewski, T. Polymer. 1999, *40*, 2347–2365.
29. Schönherr, H.; Waymouth, R.M.; Hawker, C.J.; Frank, C.W. ACS Polym. Mater. Sci. Eng. 2001, *84*, 453–454.
30. Sakai, Y.; Imai, M.; Kaji, K.; Tsuji, M. J. Cryst. Growth 1999, *203*, 244–254.
31. Bassett, D.C.; Vaughan, A.S. Polymer. 1985, *26*, 717–725.
32. Vaughan, A.S.; Bassett, D.C. Polymer. 1988, *29*, 1397–1401.
33. Sutton, S.J.; Izumi, K.; Miyaji, H.; Kukao, F.; Miyamoto, Y. Polymer. 1996, *37*, 5529–5532.
34. Izumi, K.; Ping, G.; Hashimoto, M.; Toda, A.; Miyaji, H.; Miyamoto, Y.; Nakagawa, Y. In *Advances in the Understanding of Crystal Growth Mechanisms*;

Nishinaga, T., Nishioka, K., Harada, J., Sasaki, A., Takei, H., Eds.; Elsevier Science: Amsterdam, 1997; 337–348.
35. Sutton, S.J.; Izumi, K.; Miyaji, H.; Miyamoto, Y.; Miyshita, S. J. Mater. Sci. 1997, *32*, 5621–5627.
36. Sawamura, S.; Miyaji, H.; Izumi, K.; Sutton, S.J.; Miyamoto, Y. J. Phys. Soc. Jpn. 1998, *67*, 3338–3344.
37. Taguchi, K.; Miyaji, H.; Izumi, K.; Hoshino, A.; Miyamoto, Y.; Kokawa, R. Polymer. 2001, *42*, 7443–7447.
38. Taguchi, K.; Miyaji, H.; Izumi, K.; Hoshino, A.; Miyamoto, Y.; Kokawa, R. J. Macromol. Sci. Phys. B 2002, *41*, 1033–1042.
39. Hobbs, J.K.; McMaster, T.J.; Miles, M.J.; Barham, P.J. Polymer. 1998, *39*, 2437–2446.
40. Kikkawa, Y.; Abe, H.; Iwata, T.; Inoue, Y.; Doi, Y. Biomacromolecules 2001, *2*, 940–945.
41. Reiter, G.; Sommer, J.-U. J. Chem. Phys. 2000, *112*, 4376–4383.
42. Wang, M.; Braun, H.-G.; Meyer, E. Macromol. Rapid. Commun. 2002, *23*, 853–858.
43. Barabási, A.-L.; Stanley, H.E. *Fractal Concepts in Surface Growth*; Cambridge University Press: New York, 1995.
44. Ben-Jacob, E.; Deutscher, G.; Garik, P.; Goldenfeld, N.D.; Lareah, Y. Phys. Rev. Lett. 1986, *57*, 1903–1906.
45. Hobbs, J.K.; Humphris, A.D.L.; Miles, M.J. Macromolecules 2001, *34*, 5508–5519.
46. Li, L.; Chan, C.-H.; Li, J.-X.; Ng, K.-M.; Yeung, K.L.; Weng, L.-T. Macromolecules 1999, *32*, 8240–8242.
47. Li, L.; Chan, C.-H.; Yeung, K.L.; Li, J.-X.; Ng, K.-M.; Lei, Y.-G. Macromolecules 2001, *34*, 316–325.
48. Zhou, W.; Cheng, S.Z.D.; Putthanarat, S.; Eby, R.K.; Reneker, D.H.; Lotz, B.; Magonov, S.; Hsieh, E.T.; Geerts, R.G.; Palackal, S.J.; Hawley, G.R.; Welch, M.B. Macromolecules 2000, *33*, 6861–6868.
49. Koleske, J.V.; Lundberg, R.D. J. Polym. Sci. 1969, *7*, 795–807.
50. Ong, C.J.; Price, F.P. J. Polym. Sci. Polym. Symp. 1978, *63*, 45–58.
51. Ong, C.J.; Price, F.P. J. Polym. Sci. Polym. Symp. 1978, *63*, 59–75.
52. Mareau, V.H.; Prud'homme, R.E. Macromolecules 2003, *36*, 675–684.
53. Mareau, V.H.; Prud'homme, R.E. Macromolecules 2002, *35*, 5338–5341.
54. Srinivas, S.; Babu, J.R.; Riffle, J.S.; Wilkes, G.L. J. Macromol. Sci. Phys. B 1997, *36*, 455–474.
55. Singh, S. PhD dissertation, Studies of the crystallization kinetics and morphology of poly(dimethysiloxane), McGill University, 2002.
56. Clark, M.B., Jr.; Burkhardt, C.A.; Gardella, J.A., Jr. Macromolecules 1989, *22*, 4495–4501.
57. De Cupere, V.M.; Rouxhet, P.D. Polymer. 2002, *43*, 5571–5576.
58. Hayes, N.W.; Beamsou, G.; Clark, D.T.; Law, D.S.-L.; Raval, R. Surface Interf. Anal. 1996, *24*, 723–728.
59. Bates, F.S.; Fredrickson, G.H. Annu. Rev. Phys. Chem. 1990, *41*, 525–557.

60. Weimann, P.A.; Hajduk, D.A.; Chu, C.; Chaffin, K.A.; Brodil, J.C.; Bates, F.S. J. Polym. Sci. Polym. Phys. 1999, *37*, 2053–2068.
61. Zhu, L.; Cheng, S.Z.D.; Calhoun, B.H.; Ge, Q.; Quirk, R.P.; Thomas, E.L.; Hsiao, B.S.; Yeh, F.; Lotz, B. Polymer. 2001, *42*, 5829–5839.
62. Zhu, L.; Calhoun, B.H.; Ge, Q.; Quirk, R.P.; Cheng, S.Z.D.; Thomas, E.L.; Hsiao, B.S.; Yeh, F.; Liu, L.; Lotz, B. Macromolecules 2001, *34*, 1244–1251.
63. Hong, S.; Macknight, W.J.; Russell, T.P.; Gido, S.P. Macromolecules 2001, *34*, 2876–2883.
64. Cormia, R.L.; Price, F.P.; Turnbull, D. J. Chem. Phys. 1962, *37*, 1333–1340.
65. Santana, O.O.; Müller, A. J. Polym. Bull. 1994, *32*, 471–477.
66. Jiang, S.; Qioa, C.; Tian, S.; Ji, X.; An, L.; Jiang, B. Polymer. 2001, *42*, 5755–5761.
67. Giannelis, E.P.; Krishnamoorti, R.; Manias, E. Adv. Polym. Sci. 1999, *138*, 107–147.
68. Ogata, N.; Kawakage, S.; Ogihara, T. Polymer. 1997, *38*, 5115–5118.

3

Deformation, Stretching, and Relaxation of Single-Polymer Chains: Fundamentals and Examples

GARY W. SLATER, YANNICK GRATTON, MARTIN KENWARD, LAURETTE MCCORMICK, and FRÉDÉRIC TESSIER University of Ottawa, Ottawa, Ontario, Canada

I. INTRODUCTION

Numerous practical problems involve the deformation (and the subsequent relaxation) of single polymer chains by external factors (forces, fields, confinement, etc.) [1]. Although this fully justifies why scientists study the statistical properties of single-polymer chains under various constraints, such investigations are also important for fundamental reasons. First, single-molecule studies represent a strong test of our understanding of basic physics; for instance, we can access microscopic information such as molecular conformations and trajectories, not only ensemble averaged values. Second the simplicity of such problems suggests that it should be possible to develop models with fewer approximations, thus leading to exact solutions. Third, current computing power makes it possible to simulate many single-molecule problems with atomistic or united-atom molecular dynamics algorithms [2–5]. Finally, current experimental techniques such a single-molecule fluorescence microscopy allow us to visualize single-polymer molecules for long periods of time, thus allowing us to test our models and simulations in great detail [6]. It should also be mentioned that our understanding of many-chain problems implicitly relies on our knowledge of the behavior of single-polymer chains in various environments [7–10].

In this chapter, we present some basic theoretical ideas as well as several problems related to deformed polymers in good solvents. The selection of topics is obviously based on the personal interests of the authors. We begin with a short summary of the equilibrium properties of single polymers. Section II then describes the physics of polymer chains stretched by external

forces. In Sections III and IV, we look at the role of polymer deformation on the sieving properties of new microfluidic devices. A related problem, that of tethered polymer chains stretched by strong flows, is examined in Section V. In some cases, free polyelectrolytes may stretch during electrophoresis because of subtle electromechanical effects; we present such an example in Section VI.

An isolated polymer placed in a good solvent, subjected to Brownian motion induced by the surrounding fluid, behaves like a three-dimensional random walk (RW) whose nodes move randomly while keeping the connectivity of the chain. In reality, however, backbone rigidity (BR), as well as excluded volume (EVI) and hydrodynamic (HI) interactions, often modify this simplistic picture in nontrivial ways. Nevertheless, because RW models can be solved exactly, they still play a major role in polymer science. Our approach in this section will thus be to introduce RW concepts first and then to discuss the impact of nonideal factors such as BR, EVI, and HI.

A. Static Properties of Polymer Chains

The freely jointed chain (FJC), the simplest possible model of a polymer (Fig. 1a), is a series of N links of fixed length b (the Kuhn length) connected by perfect joints [9]. The mean square end-to-end distance $\langle h^2 \rangle$ of a RW is given by the exact result

$$\langle h^2 \rangle = Nb^2 \tag{1}$$

For long chains, the probability distribution function $p(h)$ becomes Gaussian:

$$p(h) \sim \exp\left(-\frac{3h^2}{2\langle h^2 \rangle}\right) = \exp\left(-\frac{3h^2}{2Nb^2}\right), \quad N \gg 1, \; h \ll Nb \tag{2}$$

The statistical properties of the end-to-end vector $\boldsymbol{h}$ are thus identical to those of a "dumbbell" consisting of two beads connected by a harmonic spring of strength:

$$K = \frac{3k_B T}{Nb^2} = \frac{3k_B T}{\langle h^2 \rangle} \tag{3}$$

Indeed, the spring potential energy $U = 1/2\ Kh^2$ then leads to a Boltzmann distribution given precisely by Eq. (2). This is called an "entropic spring" because $\ln(p(h)) \propto K$ is related to the entropy of polymer conformations with an end-to-end distance h. For a given value of h, the entropic force keeping the chain ends together is thus given by

$$f(h) = \frac{\partial(\frac{1}{2}\, Kh^2)}{\partial h} = \frac{3k_B T}{Nb^2} h \tag{4}$$

This indicates that the equilibrium length of the spring is zero.

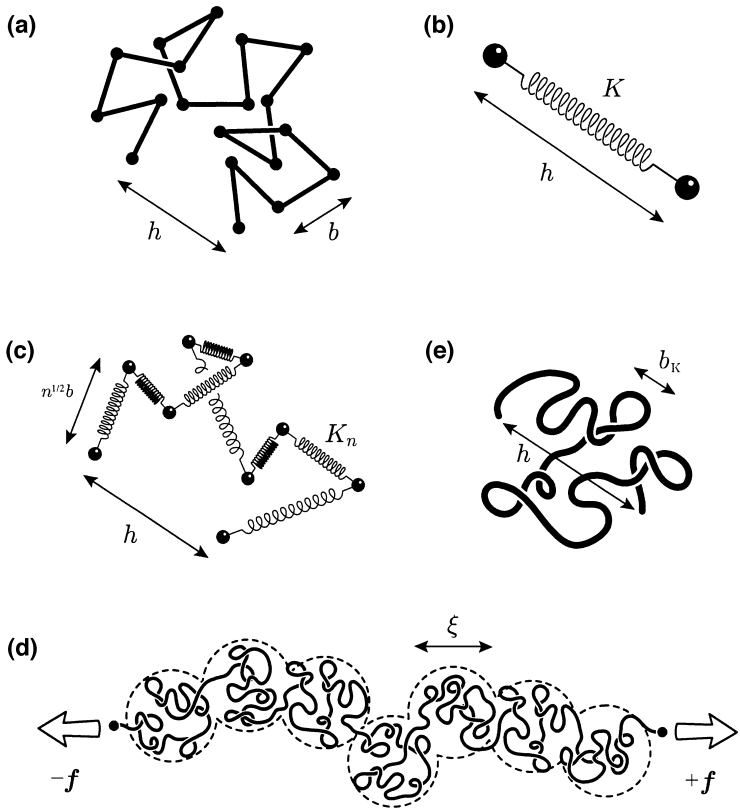


Figure 1 (a) The freely jointed chain (FJC) is made of N links of fixed length b. Because there is no constraint on the angle between two consecutive links, we have $\langle h^2 \rangle = Nb^2$. (b) The dumbbell: two beads are connected by an entropic spring with a spring constant $K = 3k_BT/Nb^2$, giving it a mean thermodynamic extension $\langle h^2 \rangle = Nb^2$. The spring replaces the N links of the FJC. (c) The Rouse chain: groups of n links are replaced by a spring (K_n) and a bead. The bead–spring chain is made of M springs and $M + 1$ beads, and we still have $\langle h^2 \rangle = Nb^2$. (d) The Pincus blobs for a stretched chain. (e) A wormlike chain is characterized by the Kuhn length b_K that plays the role of the segment size b in the FJC.

Although long RW chains can be replaced by two beads and an entropic spring (Fig. 1b), such a dumbbell model cannot describe short length scales. Because RWs are additive, we can use $M > 1$ springs and $M + 1$ beads instead of a dumbbell (Fig. 1c). We thus replace $n < N$ consecutive links of length b by an entropic spring (of mean length $n^{1/2}b$) with an elastic constant $K_n = 3k_BT/nb^2$. The end-to-end distance of the resulting chain of $M + 1$ "beads" and $M = N/n$ springs is conserved: $\langle h^2 \rangle = M(nb^2) = Nb^2$. This bead–spring (or Rouse) chain has the same statistical properties as the FJC. However, the Rouse chain also allows one to calculate dynamical properties.

B. Dynamics of Polymer Chains

A polymer chain in a good solvent continuously changes conformation in response to random Brownian forces. The equilibrium dynamics of a Rouse chain (Fig. 1c) can be analyzed in terms of relaxation (normal) modes. With harmonic springs, we can actually decouple these modes. If $R \equiv \{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_N\}$ are the positions of the N Rouse beads, the Langevin equation of motion can be written as (in matrix form)

$$\zeta \frac{\partial R(t)}{\partial t} = -KAR + \Phi(t) \tag{5}$$

The forces acting on the beads are the friction with the solvent (ζ is the bead's friction coefficient), the forces due to the entropic springs connecting the beads ($-KAR$), and, finally, the stochastic force $\Phi(t)$ due to the collisions with the solvent molecules (these forces have a zero-mean Gaussian distribution). Inertia is neglected because we are in a viscous fluid. This system of coupled harmonic oscillators is fully defined by the $N \times N$ force matrix A. For a linear chain, we can diagnolize the tridiagonal matrix A to obtain the frequencies and amplitudes of the modes. The eigenvalues of A are given by

$$\lambda_i = 4 \sin^2\left(\frac{i\pi}{2N}\right) \cong \left(\frac{i\pi}{N}\right)^2, \quad i = 0, 1, 2, \ldots, N-1 \tag{6}$$

where the last result applies when $i \ll N$. Eigenmode i ($i \geq 1$) describes chain relaxation on an intermediate length scale that corresponds to the relative motion of chain "blobs" containing N/i Rouse elements. The translation of the center of mass (the zeroth mode) predicts that the diffusion coefficient should scale like $D \sim 1/N$. Note that this result can also be derived from the Central Limit Theorem because the beads of the Rouse chain are hydrodynamically independent—such a chain is called a free-draining chain.

Of particular interest is the relaxation of the chain as described by the decay of the autocorrelation function for the end-to-end vector $\boldsymbol{h}(t)$:

$$C(t) \equiv \frac{(\boldsymbol{h}(t) \cdot \boldsymbol{h}(0))}{\langle h^2 \rangle} = \frac{8}{\pi^2} \sum_{i}^{c} \frac{1}{i^2} \exp\left(-\frac{t}{\tau_i}\right) \tag{7}$$

The global equilibrium conformation of a Rouse chain thus relaxes (i.e., looses memory of its initial state) following a sum of exponentials with characteristic times

$$\tau_i = \frac{\zeta}{\lambda_i K} \cong \frac{1}{i^2} \frac{\zeta N^2 b^2}{3\pi^2 k_B T} \equiv \frac{1}{i^2} \tau_1 \tag{8}$$

The long-time ($t > \tau_3 = \tau_1/9$) decay is described by the *terminal* relaxation time $\tau_1 \sim N^2$.

We recently examined the relaxation of Rouse chains with random spring constants [11]. Although such disorder can lead to broad distributions of relaxation times, the relaxation spectrum retains well-separated modes. Therefore, random chains still decay according to a single exponential at long times. However, the relaxation of ensembles of random chains is better described by a stretched exponential:

$$C(t) = \exp\left(-\left[\frac{t}{\tau}\right]^{\beta}\right) \tag{9}$$

where the stretched exponent $0 < \beta \leq 1$ decreases with the amount of disorder. Such behavior has been reported for complex disordered systems, including randomly branched polymers [12].

C. Real Chains and Real Solvents

Adding HI and EVI to the Rouse model is more realistic, but it leads to models that cannot be solved exactly. One must then rely on scaling arguments [7]. The inclusion of EVI leads to chain swelling and Eq. (1) then becomes

$$\langle h^2 \rangle \cong N^{2\nu} b^2 \tag{10}$$

Here, Flory's exponent $\nu \cong 3/5$ replaces the RW value $\nu = 1/2$ and b is an effective segment size. The swelling driven by EVI is more important for longer-polymer chains. A direct impact of this swelling is the reduction of the entropic spring constant now given by

$$K \cong \frac{3k_B T}{(h^2)} \sim \frac{1}{N^{2\nu}} \sim \frac{1}{N^{6/5}} \tag{11}$$

EVIs reduce the entropy of swollen chains, thus softening the entropic springs.

Hydrodynamic interactions affect only the dynamics of a chain. Zimm showed that HIs transform the free-draining chain into a hydrodynamically impermeable coil with a friction coefficient given by [8]

$$\zeta_H = 6\pi\eta R_H \tag{12}$$

where η is the solvent's viscosity and $R_H \sim \langle h^2 \rangle^{1/2} \sim N^{\nu}$ is the chain's hydrodynamic radius. Furthermore, the diffusion coefficient now scales like $D = k_B T/\zeta_H \sim N^{-\nu}$. The scaling relation $D\tau \sim R_H^2$ then leads to the scaling of the chain terminal relaxation time:

$$\tau_1 \sim \frac{R_H^2}{D} \sim R_H^3 \sim \langle h^2 \rangle^{3/2} \sim N^{3\nu} \tag{13}$$

These changes to the various laws actually introduce qualitatively different physical effects in many situations of practical interest [7].

II. PULLING THE ENDS: THE FINITE EXTENSIBILITY OF POLYMER CHAINS

Although a FJC cannot be stretched beyond its maximum extension $h_{\max} = Nb$, harmonic springs do not have such a limitation. Therefore, the spring–bead models described in Sections I.A and I.B cannot be used to study the effect of extensional forces on FJCs (unless we remain in the $h \ll h_{\max}$ range). For a more realistic physical picture, a nonharmonic spring with finite extensibility is required; that is, we need an entropic spring potential that diverges when $h = h_{\max}$. Such springs are called finitely extensible nonlinear elastic, or FENE. Unfortunately, most FENE potentials actually change the exact mean value $\langle h^2 \rangle = Nb^2$. In this section, we derive FJC FENE potentials in the stress and strain ensembles. We also examine the stretching of real chains by external forces.

A. Stretching an Ideal FJC: The Stress Ensemble

Here, fixed forces f are applied to the chain ends and the end-to-end distance h fluctuates (h and f are conjugate thermodynamic variables, much like P and V, or H and M). We thus want to compute the mean value $\langle h(f) \rangle$. Each FJC

segment is like a dipole forced to orient in the direction of an external field. The exact solution to this problem is well known [13]:

$$\frac{\langle h(F)\rangle}{h_{\max}} = L(F) \equiv \coth(F) - \frac{1}{F}$$

$$\cong \begin{cases} (1/3)F - (1/45)F^3 + (2/945)F^5 - O(F^7), & F \ll 1 \\ 1 - \dfrac{1}{F}, & F \gg 1 \end{cases} \tag{14}$$

where $F = fb/k_BT$ is the scaled force and $L(F)$ the Langevin function. As expected, the harmonic result in Eq. (4) is the first term, $F/3$, of the $F \ll 1$ expansion, and $\langle h\rangle \to h_{\max}$ when $F \to \infty$. Because the Langevin function cannot be inverted analytically, we must invert the series expansions to obtain a replacement for Eq. (4):

$$F(H) = L^{-1}(H) \cong \begin{cases} 3H + (9/5)H^3 - (297/175)H^5 + O(H^7), & H \ll 1 \\ \dfrac{1}{1-H}, & 1 - H \ll 1 \end{cases} \tag{15}$$

where $H = \langle h\rangle / h_{\max}$ is the scaled chain extension. Simple interpolation formulas based on Eq. (15) can be constructed using the ratio of two polynomials; the simplest one is [14]

$$F(H) \cong \frac{3H - H^3}{1 - H^2} \tag{16}$$

Strangely, however, the Warner approximation [14] $F(H) = 3H/(1-H^2)$ is more widely used even though it overestimates the force by 50% close to full extension.

The stiffer FENE potential correctly limits the maximum spring extension to $h_{\max} = Nb$, but it also reduces the equilibrium spring extension by a factor $\cong N/(N+1)$. Therefore, replacing a FJC by a Rouse chain that uses the FENE force defined by Eq. (14) does not conserve the size $\langle h^2\rangle$ of the original molecule. This is because the stress ensemble actually describes the mean chain extension $\langle h(f)\rangle$ for a fixed external force f. In the Rouse model of polymer dynamics, on the other hand, we need to know the mean force $\langle f(h)\rangle$ pulling together the ends of an entropic spring whose instantaneous

end-to-end distance is h. This problem is, in fact, related to the strain ensemble.

B. Stretching an Ideal FJC: The Strain Ensemble

Calculating the mean force $\langle f(h) \rangle$ given a fixed end-to-end distance h is quite a complicated problem. Although Treloar did derive an exact solution [13], the latter is not analytic and cannot easily be used. More recently, Glatting et al. [15] developed an alternative (and analytic) approximate expression with the following limits:

$$F(H) \cong \begin{cases} \left(3 - \dfrac{3}{N}\right)H + \left(\dfrac{9}{5} - \dfrac{99}{25N}\right)H^3 + \left(\dfrac{297}{175} - \dfrac{837}{175N}\right)H^5 + O(H^7), & H \ll 1 \\ \dfrac{1}{NH} + \dfrac{N-2}{N(1-H)}, & 1 - H \ll 1 \end{cases} \tag{17}$$

Thus, the two ensembles simply differ by factors of order $1/N$ [compare Eqs. (15) and (17)] and thus become identical in the thermodynamic limit $N \gg 1$, as expected. The simplest interpolation formula is now

$$F(H) \cong \frac{3(1 - 1/N)H - (1 + 1/N)H^3}{1 - H^2} \tag{18}$$

This expression, which gives the leading terms of Eq. (17) in both limits, is quite good for the whole range of extensions $H \in [0,1]$, even for very small values of N. It is important to realize that the properties of the entropic springs now depend on N, the number of polymer segments they replace in the construction of a Rouse chain. With strain-ensemble FENE potentials, the exact relation $\langle h^2 \rangle = Nb^2$ is now satisfied to a very high degree of accuracy (the error is of order $1/N^2$). In conclusion, in order to replace a given FJC by a series of beads and springs that has the proper maximum extension and mean square end-to-end distance, we need a strain-ensemble FENE potential.

C. Stretching a Real Chain: Scaling Arguments

In the absence of exact solutions for EVI chains, we will rely on scaling arguments to study a real chain being stretched by constant forces $\pm \boldsymbol{f}$ applied at both ends—the stress ensemble. We can actually predict the existence of four different regimes here. In the low-force limit, the spring constant (11) gives the linear regime (using $f \sim Kh$)

$$\text{I}: \qquad \frac{\langle h \rangle}{b} = \frac{1}{b}\frac{f}{K} \cong \frac{N^{2\nu}b}{k_B T} f \cong N^{2\nu} F \sim N^{6/5} f^1, \quad F < N^{-\nu} \tag{19}$$

The linear regime applies only to small deformations $\langle h \rangle < N^{\nu} b$. For a given scaled force F, Eq. (19) is thus limited to chains with $N < g \cong F^{-1/\nu}$ segments; the corresponding critical coil size is $\xi = \langle h(g) \rangle = g^{\nu} b \cong b/F$. The so-called Pincus regime [16] occurs when $\xi < \langle h \rangle \ll Nb$. The polymer conformation is then an ideal chain of N/g aligned "Pincus blobs" of size ξ (Fig. 1d), and the chain conformation inside each blob is essentially unperturbed by the external force. The extension is given by

$$\text{II}: \qquad \frac{\langle h \rangle}{b} \cong \frac{N}{g}\frac{\xi}{b} \cong NF^{(1/\nu)-1} \sim N^1 f^{2/3}, \quad N^{-\nu} < F < 1 \tag{20}$$

The change from $\langle h \rangle \sim f^1$ to $\langle h \rangle \sim f^{2/3}$ is quite radical.

Excluded volume interactions can be neglected when the chain is highly stretched. This occurs roughly when $F \cong 1$. In this regime, one is back to a FJC, and we should recover Eq. (14):

$$\text{III}: \qquad \frac{\langle h \rangle}{b} \cong N\left(\coth(F) - \frac{1}{F}\right) \cong N^1\left(1 - \frac{1}{F}\right), \quad F > 1 \tag{21}$$

Finally, the last regime (IV) is found when the force is large enough to disturb the chemical bonds, such that we may get $h > Nb$. The behavior of the chain is then nonuniversal; that is, it is related to its chemical structure. Beyond a certain critical force, one may even observe chain scission. These effects are beyond the scope of this chapter.

D. Wormlike Chains

In this section, we neglect EVIs and, instead, explore the effects of finite rigidity. Indeed, real polymers possess nonzero bending rigidity (BR), which is usually described in terms of a Kuhn length b_K [9]. If L_c is the contour length of such a semiflexible polymer chain (Fig. 1e), Eq. (1) can be rewritten as $\langle h^2 \rangle = bL_c$, with $L_c = Nb$. With BR, the polymer swells and this relation is no longer valid. The Kuhn length of a polymer with BR is defined as the length which allows us to replace Eq. (1) by

$$\langle h^2 \rangle = b_K L_c \tag{22}$$

for very long polymer chains ($L_c \gg b_K$). If we further define $N_K \equiv L_c/b_K$ as the size of the polymer in terms of Kuhn lengths, we recover the equivalent of Eq. (1): $\langle h^2 \rangle = N_K b_K^2$. In other words, a long semiflexible chain behaves like a normal FJC chain if we replace the real segment (or monomer) size by the Kuhn length b_K.

For shorter chains, however, Eq. (22) is inadequate. A detailed calculation, based on the theory of wormlike chains (WLC) actually gives the following expression [9]:

$$\langle h^2 \rangle = b_K \left[L_c + \frac{b_K}{2} (e^{-2L_c/b_K} - 1) \right] \cong \begin{cases} L_c^2, & L_c \ll b_K \\ b_K L_c, & L_c \gg b_K \end{cases} \tag{23}$$

The properties of semiflexible chains are still poorly understood. Because DNA is very rigid ($b_K \cong 100$ nm) and has been extensively studied using, for example, optical tweezers and atomic force microscopes (AFMs), there has been a growing interest in the theory of WLCs over the last decade. Of central interest is the force–extension relation for semiflexible chains in the stress ensemble. The presence of a bending potential changes the physics in a fundamental way and Eq. (15) is no longer applicable. The most used WLC force–extension law is the interpolation formula derived by Marko and Siggia [17]:

$$F_K(H) = 2H - \frac{1}{2} + \frac{1}{2(1-H)^2} \cong \begin{cases} 3H + O(H^2), & H \ll 1 \\ \dfrac{1}{2(1-H)^2}, & 1 - H \gg 1 \end{cases} \tag{24}$$

where the scaled force is now given by $F_K = f b_K / k_B T$ because b_K is the relevant segment length for a semiflexible chain. We note that, to first order, the weak force result $F_K \cong 3H$ is identical to the one found for the FJC [see Eq. (15)]. The stretching process actually follows two steps. First, the effective segments orient along the direction dictated by the external force, and the semiflexible nature of the chain plays no direct role (besides a trivial redefinition of the segment length b_K). However, when $F_K > 1$, the force is large enough to affect the molecular conformation on scales smaller than the segment size b_K. In this second stage, the force begins to stretch the semiflexible effective segments themselves. For instance, we note that Eq. (23) predicts that at equilibrium, the effective segments, which have a contour length b_K, possess a root-mean-square end-to-end distance of only $\cong 3/4\, b_K$. This second stage thus requires additional energy. A useful way to compare these results is to look at the asymptotic behavior:

$$1 - H \sim \begin{cases} \dfrac{1}{F} & \text{(FJC)} \\ \dfrac{1}{\sqrt{F_K}} & \text{(WLC)} \end{cases} \tag{25}$$

Close to full extension, $H = 1$, the two models thus predict different power laws. For a given value of H, the force required to stretch a WLC is larger than

that required to stretch a FJC. Experimental data for DNA [1] clearly indicates that indeed the FJC formula (15) underestimates the force and that the WLC law, $1 - H \sim 1\sqrt{F_K}$, applies.

More recently, Winkler [18] proposed a theory for semiflexible chains of Gaussian segments. For long chains ($L_c \gg b_K$), his work suggests the alternative force–extension law:

$$F_K(H) = \frac{3H}{(1-H^2)^2} \cong \begin{cases} 3H + 6H^3 + O(H^5), & H \ll 1 \\ \dfrac{3}{4(1-H)^2}, & 1 - H \ll 1 \end{cases} \tag{26}$$

This result is slightly different from the Marko and Siggia formula, but it does agree with it in the two asymptotic limits. In particular, the $1 - H \sim 1\sqrt{F_K}$ law still applies. In the short-chain limit ($L_c \ll b_K$), on the other hand, Winkler predicts

$$F_K(H) = \frac{3H(1 + 1/N_K)}{1 - H^2} \tag{27}$$

Interestingly, this is essentially Eq. (16). We note that stretching the Kuhn segments to full extension also requires a force that scales like $F_K \sim 1/(1 - H)$.

III. POLYMER STRETCHING OVER POTENTIAL BARRIERS

The theoretical concepts introduced earlier have significant practical applications. Indeed, the deformation of single macromolecules can often improve—or in some cases lie at the very heart of—the operation of analytical devices. In particular, recent leaps in the designability of fluidic systems at the microscopic or even nanoscopic level offer unprecedented opportunities to modulate the conformation of individual macromolecules by means of molecular-scale obstacles and confinement. The induced conformational changes give rise to significant "entropic forces" which can be brought to serve a useful purpose, e.g., sorting biopolymers according to size or topology (linear chain, ring, knot, etc.). The perspective adopted here is slightly different compared to the previous section, in that we are not actively deforming the polymer by pulling on its ends. Rather, we consider how a polymer deforms when it encounters a potential barrier and how this deformation, in turn, affects how the polymer may escape from the potential well.

A striking example of the relevance of conformational entropy for the design of analytical tools is the *entropic trap* array built by the Craighead group [19–23]. Schematically represented in Fig. 2a, it consists of a micro-

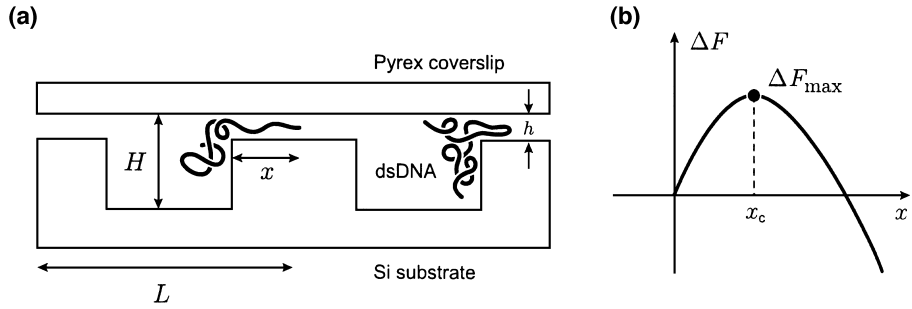


Figure 2 (a) The microfluidic channel built by the Craighead research group to separate double-stranded DNA (dsDNA) according to size. In the actual devices, $H \approx 1$ μm, $h \approx 0.1$ μm, and the period L can range from a few microns to a few tens of microns. The molecules are pulled toward the right by an external electric field, but their migration is hindered by the shallow constrictions. (b) A sketch of the change in the free energy ΔF of the molecules as a function of the hernia length x. The effective activation energy corresponds to the maximum of ΔF, which occurs at a critical hernia length x_c.

fluidic channel with alternating deep wells, where the polymer molecule sits comfortably, and shallow constrictions, which the molecule can only cross by adopting an elongated shape. Elongation incurs a large entropy cost; thus, at equilibrium, the molecule is in fact confined to a given well (hence, the name *entropic trap*). With an external force oriented along the channel direction, however, the molecule is encouraged to travel from one trap to the next by slithering across the constriction. The rate at which this *translocation* process occurs is determined by the balance between the external and the entropic forces.

The prevailing model for this system simply considers scaling relations for the change in free energy ΔF when a coil section of length x—a hernia—stretches out inside the constriction. For the case at hand, an electric field of strength E is pulling on a DNA strand (a polyelectrolyte) inside the channel, so we have $\Delta F_{\text{entropic}} \sim x$ (because conformational entropy is an extensive molecular variable) and $\Delta F_{\text{electric}} \sim -x^2 E$ (because both the charge and the span of the hernia are proportional to x). Hence, the total change in the free energy of the polymer as a function of hernia size follows

$$\Delta F \sim x - x^2 E \tag{28}$$

Equation (28) reveals the presence of a potential barrier (see Fig. 2b), and differentiation predicts that both the critical hernia length x_c and the cor-

responding barrier height $\Delta F_{\max}$ scale like $1/E$. The net rate of crossing, κ, thus takes the Arrhenius form

$$\kappa = \kappa_0 e^{-\Delta F_{\max}/k_B T} = \kappa_0 e^{-E_0/E} \tag{29}$$

where κ_0 and $E_0 \sim 1/k_B T$ are independent of the field intensity E. In Eq. (29), there is no explicit dependence of the rate κ on the total number of monomers N because we have overlooked the three-dimensional aspect of the system (going *into* the page in Fig. 2a). In reality, larger coils expose more monomers to the constriction than smaller ones, thus generating more herniae per unit time and therefore increasing their chances of escaping from the well. This contribution turns up as an N dependence of the prefactor κ_0, making size separation possible. Furthermore, the field E pushes the coil against the constriction, which enhances this effect. Note that although they lose more entropy in the narrow channel, longer molecules move faster here!

The above-discussed system is of course but one example of a microdevice intended to manipulate and analyze macromolecules, and recent literature abounds in clever designs that exploit conformational behavior (stretching and/or relaxation) of single molecules. The kind of model presented earlier is quite generic and thus generally helpful to describe the operation of devices that rely on similar principles. However, it is possible to further refine our understanding by considering not only the dynamics of the polymer chain as a whole, but of each of its segments explicitly. The fundamental question here is: *How does a polymer chain cross a potential barrier?* This is a generalization of the well-known Kramers problem for Brownian particles to the case of stringlike objects, and Sebastian and Paul have published instructive accounts of the solution for the case of long chains (i.e., chains with a contour length much larger than the barrier width) [24–28]. The mathematical derivation of the translocation rates and the total crossing time are rather involved, extending beyond the scope of this short account, so we only review here the simple calculation of the activation energy.

The approach is based on the Rouse model [8], introduced in Section I, considered here in the continuum limit (i.e., we treat the bead index n as a continuous variable ranging from 0 to N), a reasonable approximation for long chains. The polymer is placed in a potential double well, such as the one shown in Fig. 3, and we investigate how it may overcome the potential barrier. Taking $\boldsymbol{R}(n, t)$ as the position of bead n at time t, and $V(\boldsymbol{R}(n, t))$ as its potential energy, we can write Newton's equation of motion for the (massless) beads:

$$\zeta \frac{\partial \boldsymbol{R}(n,t)}{\partial t} = K \frac{\partial^2 \boldsymbol{R}(n,t)}{\partial n^2} - \nabla V(\boldsymbol{R}(n,t)) + \Phi(n,t) \tag{30}$$

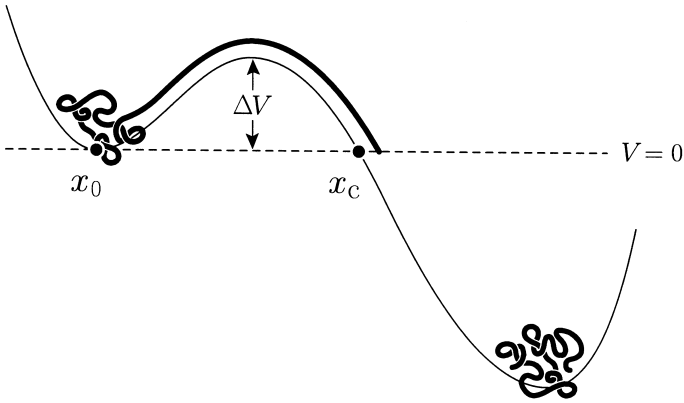


Figure 3 Schematic representation of a polymer coil crossing a potential barrier. The coil is initially localized around the potential minimum at x_0 and initiates translocation by stretching over the potential barrier ΔV up to a point x_c where $V(x_c) = V(x_0)$.

where ζ is the friction coefficient of a bead and $\Phi(n,t)$ is a random force. The spring constant K accounts for entropic effects; as earlier, $K = 3k_BT/\langle h^2 \rangle$, with $\langle h^2 \rangle$ the mean square length of the springs. In addition, we impose the boundary conditions $(\partial R/\partial n)|_{n=0} = (\partial R/\partial n)|_{n=N} = 0$ to reflect the fact that the two ends of the chain are free [8].

For our purpose, it suffices to restrict the analysis to a one-dimensional case and to the deterministic part of Eq. (30) (i.e., $\boldsymbol{R} = \boldsymbol{x}$ and $\boldsymbol{\Phi} = 0$). We are primarily interested in the situations of static equilibrium, where the position of the beads are independent of time. In this case, the equation of motion reduces to

$$K \frac{d^2x(n)}{dn^2} = \frac{dV(x(n))}{dx} \tag{31}$$

An obvious solution of this equation is when the beads are localized in one potential well; both sides of Eq. (31) then, vanish. However, there is another solution where a portion of the chain just straddles the potential barrier, as depicted in Fig. 3. The total free energy of the chain is then most interesting because it gives the activation energy for the onset of translocation (i.e., the effective barrier height seen by the polymer). Indeed, following Sebastian and Paul [24], we notice that Eq. (31) is, in fact, Newton's equation of motion for a particle of mass K moving in a potential $-V$, where n plays

the role of time. By analogy with conservation of energy, we at once have that the quantity

$$U = \frac{K}{2}\left(\frac{dx}{dn}\right)^2 - V(x) \tag{32}$$

is conserved as one moves along the chain contour, from $n=0$ to $n=N$. Recalling that $dx/dn = 0$ at either end of the chain, we have $U = 0$, because the tail of the chain lies at a point where we chose $V=0$; hence, the equality

$$\frac{K}{2}\left(\frac{dx}{dn}\right)^2 = V(x) \tag{33}$$

holds everywhere along the chain. In particular, the head of the straddling portion of the chain must also lie at the point (denoted x_c in Fig. 3) where $V = 0$ on the other side of the barrier. This state of unstable equilibrium is intuitively correct when one thinks, for example, of a frictionless rope resting over a bump. To calculate the activation energy for translocation, we introduce the free-energy density along the chain contour; that is,

$$\mathscr{F}(n) = \frac{K}{2}\left(\frac{dx}{dn}\right)^2 + V \tag{34}$$

where the first term on the right is the contribution from the local stretching around bead n, and the second term is simply the potential energy of bead n. The total free energy of the chain may then be expressed as

$$F = \int_0^N \mathscr{F}(n)\, dn \tag{35}$$

Combining Eqs. (33)–(35) and assuming that $x(n)$ is an increasing function of n, we find the activation free energy for the crossing process to be

$$\Delta F_a = \int_0^N 2V\, dn = \int_{x(0)}^{x(N)} 2V\left(\frac{dn}{dx}\right) dx = \int_{x_o}^{x_c} \sqrt{2KV}\, dx \tag{36}$$

This result is independent of the total length of the polymer: *The extension of the head of the chain across the barrier is sufficient to initiate translocation*. When compared with the energy $N\Delta V$ that would be required to move the N beads across the barrier *all at once*, it becomes clear that the effective height of the barrier is greatly reduced by stretching for long chains. There is yet a third solution of Eq. (31) corresponding to a chain hairpin, instead of a chain end, resting atop the potential barrier. This translocation mechanism is increasingly probable as N grows and its activation energy is easily seen to be twice that of the end-activated crossing. Such results can be applied to situations of

practical interest and, in the case of the Craighead device discussed earlier, they agree in character with Eq. (29) and with experimental data.

Park and colleagues have also published a thorough analysis of the Kramers problem for a Rouse chain, for polymer contour lengths that are not necessarily large compared to the barrier width [29–31]. They showed that the stretching induced by a convex barrier *always* reduces the activation energy for chain translocation. Therefore, the rate of crossing is always larger than in the globular limit where N beads rigidly cross the barrier simultaneously. This conformational effect is stronger for longer chains, whereas it becomes insignificant for small polymers crossing wide barriers. More interestingly, because of the competition between surmounting the barrier and stretching at the barrier top, there is a maximum in the effective barrier height at intermediate values of N; that is, there is a critical chain length at which the crossing rate is minimized.

From this short overview, we gather that molecular deformations play a determining role in the escape dynamics of polymer chains from potential wells, including those arising for entropic reasons. Qualitatively, the key point to bear in mind regarding this kind of process is that only a fraction of a long flexible macromolecule needs to overcome the potential barrier in order to initiate the directed translocation of the whole chain. The analogy with a siphon, wherein a relatively small amount of energy is required to trigger fluid to flow naturally over a gravitational potential barrier, provides a helpful picture to remember this aspect of the polymer translocation process.

IV. POLYMER–OBSTACLE COLLISIONS

As mentioned in the previous section, polymer deformation, stretching, and relaxation plays a crucial role in the separation of biological macromolecules. Many separation microdevices such as those utilized to analyze biological macromolecules (including DNA) often exploit molecular collisions between a moving analyte and various obstacles. These collisions, and the molecular deformation they generate, slow down the analyte in such a way that useful separation can be achieved. The importance of these devices in the biosciences (e.g., it is the invention of powerful DNA separation methods that allowed the Human Genome to be mapped and sequenced years ahead of schedule) has spurred a renewed interest in the dynamical behavior of single isolated polymers [32–36]. Two excellent model systems for examining single-chain dynamics and the effects of hydrodynamics are either microscopic obstacles (or arrays of obstacles) fabricated using microlithography [32] or magnetically self-assembled superparamagnetic ferrofluid particles [33,37] which can form quasiregular arrays of posts.

Electrophoresis uses an electric field to move charged analytes through a sieving system. When a DNA molecule collides with an obstacle, it generally deforms and slides around it in order to continue its migration. During the collision, the presence of both electric and mechanical forces may lead to subtle effects that will be briefly discussed in the next section. Here, we will neglect these combined effects and examine polymer–obstacle collisions from a purely mechanical point of view. Several authors have looked at different aspects of this problem because these single-polymer systems represent, in some sense, the fundamental unit of separation in such devices [32,34,36].

We can, in fact, think of two apparently analogous methods of moving a polymer through a microfluidic device. One can either apply a uniform mechanical force directly on each monomer (e.g., a sedimentation force) or, alternatively, one can use a low-Reynolds-number fluid flow with velocity v_s to force the molecule through the device. Although these approaches lead to nearly identical molecular behavior, there is a subtle difference related to the hydrodynamic properties of deformed polymer chains. It is on this interesting difference that we will focus our attention.

A. Mechanical Forces

First, consider a quiescent isotropic viscous fluid in which a polymer is moving under the influence of a uniform external field (this correponds to Fig. 4a). Each of the N monomers thus experiences an identical external force f_0 regardless of the chain conformation. The simplest such example is a rope moving in a gravitational potential. In the case of a polymer moving in a viscous liquid, this uniform external force alone is unable to deform the polymer. Without obstacles, the motion of the chain is dictated by the retarding effect of the friction between the polymer and the solvent.

The chain friction can be described by two different models: the local hydrodynamically permeable (Rouse) and nonlocal hydrodynamically impermeable (Zimm) friction models. The friction coefficient in the Rouse regime (subscript R) is given by $\xi_R = N\xi_1$, where $\xi_1 \cong 6\pi\eta b$ is the Stokes friction coefficient of a monomer of size b. Note that ξ_R is the same for all molecular conformations. However, when we include (nonlocal) hydrodynamic interactions (subscript HI), the instantaneous friction coefficient is conformationally dependent and is given by $\xi_{\mathrm{HI}} = 6\pi\eta R_H$, where R_H is the instantaneous hydrodynamic radius of the polymer. In equilibrium, Eq. (10) applies and we have $R_H \sim N^{\nu}$.

Because the friction coefficient changes with the molecular conformation in the nonlocal model, it is clear that the two models will predict different

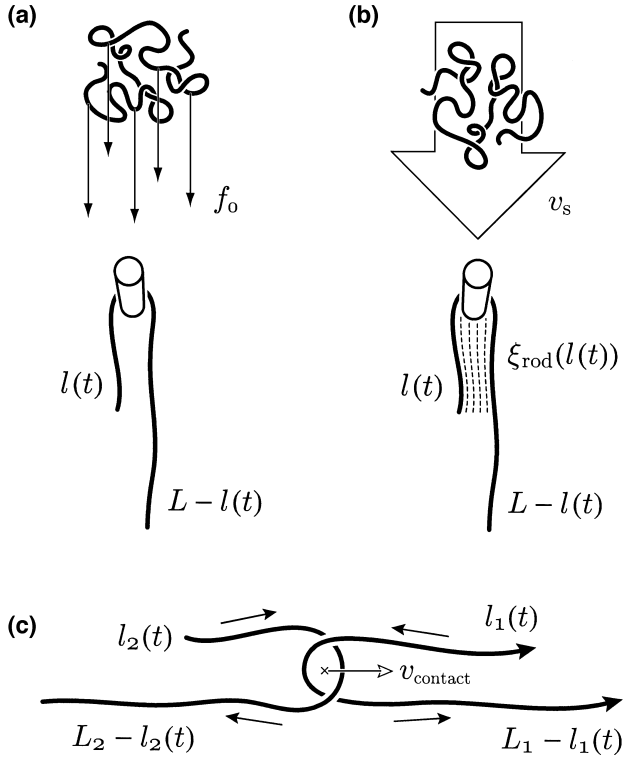


Figure 4 Schematic illustration of polymer–obstacle and polymer–polymer collisions in the external force and fluid flow regimes. The hatch marks in (b) are intended to illustrate the effective hydrodynamic coupling of the chain in the fluid flow.

collision dynamics. The question then is: How does the presence of nonlocal hydrodynamic interactions modify the retarding effect of the polymer–obstacle collisions in a sieving system?

A complete collision process comprises a series of four consecutive steps: free-flow migration, deformation, trapping release, and relaxation. First, the polymer migrates through the solvent under the influence of the applied field, with constant velocity $v_0 = Nf_0/\xi$, where ξ is the friction coefficient in either the local ($\xi = \xi_R$) or nonlocal ($\xi = \xi_{HI}$) friction model. In free-flow electrophoresis, the local model applies because the charged chain is free draining [38]; in this case, $v_0 = Nf_0/\xi = f_0/\xi_1$ is independent of the polymer size N, hence the need for sieving media. A useful collision (at time $t = 0$) will then deform the polymer, and the latter will take a pulleylike conformation

(Fig. 4), with stretched arms of length $L - l(0)$ and $l(0) \leq L/2$. The limiting case is that of central trapping, where $l(0) = L/2$. A simple approximation to the pulley formation time is to assume that the ends of the polymer travel at velocity v_0 until the pulley is fully formed. To first order, the pulley formation time is thus $\tau_p \approx (L - l(0))/v_0$. The third stage of this process (rate-determining step), the competition between the two downfield arms then begins. This competition dictates the release time of the polymer, which we denote by τ_{esc} [see Eq. (40)]. When the polymer is released from the obstacle, it is in a highly extended state. In the fourth and final stage, it relaxes to its equilibrium random coil conformation while drifting away from the obstacle, and it subsequently returns to velocity v_0. The relaxation of a chain after collision is also still not well understood [37]. If the device has periodic sets of obstacles, the polymers will undergo consecutive collisions, and the mean trapping time and probability of collision will likely depend on N, the degree of polymerization. In essence, this is a simple concept for the separation of flexible biopolymers via sieving.

In the pulleylike conformation, the polymer is no longer a random coil and is highly elongated in the field direction. For simplicity, we assume that the polymer is fully extended and that $L = L_c = Nb$ (the contour length of the polymer). This is the high-field regime, frequently used in sieving devices. More precisely, this means that the field dominates Brownian motion [36]; this occurs when $F = Nf_0 \gg k_BT/b$ [36].

In order to study the escape from the pulleylike conformation, we need to estimate the chain's friction coefficient in this state. In the Rouse picture, the friction coefficient does not change because it is independent of conformation and is given by $\xi_R^{\cap} = N\xi_1$ where the superscript $\cap$ is meant to denote the friction coefficient in the pulley (or $\cap$-shaped) conformation, although it is equal to ξ_R here. In the nonlocal case with HI, we can treat the polymer as as cylinder of length $L \backsimeq Nb$, which moves along the contour of the chain. We thus ignore the bending, the friction due to the polymer–obstacle contact, and the possible hydrodynamic coupling due to the shearlike opposing motion of the two arms. The friction coefficient of such a cylinder (from the slender body theory of macroscopic hydrodynamics [39]) is given by

$$\xi_{HI}^{\cap} = \frac{2\pi\eta L}{\ln(L/b) + \gamma^{\parallel}} \cong \frac{1}{3\ln(N)}\xi_R^{\cap} \tag{37}$$

where $\gamma^{\parallel} \cong 1/2$ is simply a numerical constant. We note that $\xi_R^{\cap} \sim \xi_{HI}^{\cap} \sim N$ to leading order. Therefore, we expect rather similar escape times.

We now analyze the equation of motion of a chain in a pulley state with an initial parameter $l(t = 0) \equiv l(0)$. At time $t \geq 0$, the driving force for the

escape of the polymer is the unbalanced length $L - 2l(t)$. The equation of motion is thus

$$\frac{dl(t)}{dt} = -\frac{f_0}{\xi^{\cap}} \frac{L - 2l(t)}{b} \tag{38}$$

and its solution is

$$l(t) = \frac{L}{2} - \left(\frac{L}{2} - l(0)\right) \exp\left(\frac{2f_0 t}{\xi^{\cap} b}\right) \tag{39}$$

where we obtain the results for the local and nonlocal models by substituting the appropriate value of the friction coefficient $\xi^{\cap}$. The escape time of the polymer can be simply obtained by solving for $t = \tau_{esc}$, at which $l(\tau_{esc}) = 0$. Subject to this constraint, we can invert Eq. (39) and solve explicitly for the escape time of the polymer:

$$\tau_{esc}(l(0)) = \ln\left(\frac{L}{L - 2l(0)}\right) \frac{\xi^{\cap} b}{f_0} \tag{40}$$

We note that this time diverges when $l(0) = L/2$. In this case, the two arms of the pulley are of equal length and the net force is zero although the net force on the obstacle is maximum. The escape time, however, is not infinite in this situation because Brownian motion quickly allows the system to get out of this unstable equlibrium state; as soon as $l(t)$ becomes non-negligible, the above-described deterministic escape process starts. Finally, we must average the previous result over all possible values of $l(0)$ between 0 and $L/2$. For a uniform distribution of initial conditions, this average gives the remarkably simple result

$$\langle \tau_{esc} \rangle = \frac{\xi^{\cap} b}{2f_0} = \frac{L}{2}\left(N\frac{f_0}{\xi^{\cap}}\right)^{-1} \tag{41}$$

We note that the escape time is that required to move over a distance $L/2$ given the effective velocity $v_0^{\cap} = Nf_0/\xi^{\cap}$, which is essentially independent of the molecular size N. We thus see that the escape time increases proportional to L in both models but that its magnitude is smaller in the model with HIs [see Eq. (37)]. However, this similarity is hiding a key difference. Indeed, whereas $v_0^{\cap} = v_0 \sim N^0$ in the Rouse model, $v_0 \sim N^{1-\nu}$ in the model with HIs. Therefore, the collisions affect the mean velocity of the molecules through the sieving device in very different ways according to these two models.

The situation with electrophoretic forces is even more complicated than this, but it has yet to be studied in detail. In short, although the polymer chain behaves like a Rouse chain between collisions, it actually behaves like a chain

deformed by a strong counterflow during the collision; this type of collision is discussed next.

B. Fluid-Induced Polymer Motion

In this section, we consider a polymer in a solvent flowing at velocity $v_s = v_0$, (see Fig. 4b). We outline how the presence of a solvent flow modifies the escape dynamics of a polymer from an obstacle. We begin by examining a centrally trapped position where $l(0) = L/2$. The force on the polymer due to the obstacle is the same as the frictional force of the two arms of the pulley imparted by the flowing solvent. The crucial qualitative difference in this example is that because the interarm distance d is much smaller than the length of the chain, we can consider the pulley to be a hydrodynamically coupled object. The entire chain can be treated as a composite object of length $L/2$ and effective diameter $d \cong b$ [36,39]. The frictional coefficient of the chain is that of a rod of length $L/2$ [i.e., $\xi_{HI}^{rod}(L/2)$ (in the central trapped position)]. The associated frictional force F_s on the chain due to the solvent flow is

$$F_s = v_s \xi_{HI}^{rod}(L/2) \tag{42}$$

This is a rather subtle point, and if, for example, the interarm spacing of the pulley were much larger (e.g., $d > L/2$), this hydrodynamic coupling would be negligible.

We now consider the general case $l(0) < L/2$. The equation of motion is a balance between the sliding motion of the chain (around the post) at speed $v_c = -dl(t)/dt$ (with a frictional resistance proportional to the total contour length L) and the driving force due to solvent flow (at velocity v_s) acting on the unbalanced length $L - 2l(t)$:

$$\xi(L)\frac{dl(t)}{dt} = -v_s \xi(L - 2l(t)) \tag{43}$$

In the limit where the polymer strands are fully stretched, the friction coefficients all scale like $\xi(x) \sim x$ and we immediately recognize that this is the same equation of motion as that which we obtained for the case of a mechanical force. Therefore, although the total force is reduced by the hydrodynamic coupling, the sliding motion is unaffected. Using Eq. (37) for $\xi_{HI}(x)$, the mean escape time is found to be given by

$$\langle \tau_{esc} \rangle = \frac{L}{2v_s} \frac{2\ln(L/b - 1)}{2\ln(L/b + 1)} \tag{44}$$

Although the effective hydrodynamic coupling affects the nature of the forces, the polymer escape time is virtually the same as that given by Eq. (41).

As pointed out by André et al. [36], the effects of hydrodynamic coupling would be most prominent in the case of a chain hooking on the obstacle with multiple loops: The frictional force would be reduced significantly due to the weak dependence of the friction coefficient on the diameter of the chain [36] and the competition between several unbalanced loops would lead to different results. This will not be treated here.

Although these simple approaches to treating the escape time of a polymer provide useful results, more detailed modelling is required in order to determine effects due to interarm interactions, shear coupling due to the relative motion of the arms of the pulley, and effects of the size of the obstacle on the escape time. Moreover, at lower field intensities, the forces are not large enough to fully stretch the two arms of the pulley and the friction coefficients must be modified accordingly. From a practical point of view, understanding how these variables affect polymer dynamics can lead to ways of optimizing separation systems. In the next subsection, we study a variant of the polymer–obstacle collision: the polymer–polymer collision.

C. Polymer–Polymer Collisions

An interesting combination of the previous two examples occurs when an external force is applied to a polymer (the polymer labeled 1 in Fig. 4c), which then collides with a free polymer (labeled 2). This collision typically results in the formation of a double pulleylike conformation as depicted in Fig. 4c. This effectively combines the two effects of a fluid flow (the relative motion of the contact point, at velocity $\upsilon_{\text{contact}}$, is, in fact, analogous to a fluid flow) and an external force on the polymer. These collisions play a crucial role in the size separation of polymers in dilute solution capillary electrophoresis [40,41]. In this sieving regime, a dilute neutral solution of polymers is used as a "mobile" sieving matrix to separate charged biomolecules such as DNA. The exact mechanism responsible for this separation is not well understood. There are currently two suspected causes for this separation. The first, proposed by Barron et al. [40,41], is that of a so-called transient entanglement mechanism, wherein the charged species collides with one or many neutral polymers, dragging them through the solvent. In turn, this slows down the motion of the chains. The other mechanism recently proposed by Ekani-Nkodo and Tinland [42], is that the dilute matrix affects the hydrodynamic flow field of the polymer, in effect increasing its hydrodynamic radius. Much work is still required in order to better understand separation in this system. In particular, it is important to understand the effects of hydrodynamics on the motion of both colliding chains.

One can immediately see that the case of a fixed post is a subset of the more general situation (seen in the two previous subsections), although the calcu-

lation of the escape of a polymer from a polymer–polymer collision is somewhat more involved than the previous two examples. Here, we outline the methodology of the calculations. Figure 4c is a schematic illustration of a polymer–polymer collision. The key phenomenological difference between this example and the previous two is that here we have a nonzero contact-point velocity v_{contact} of the two-molecule system. This has the added consequence of inducing a further frictional term due to relative motion of the whole system with respect to the stationary surrounding fluid. Moreover, compared to the case of a fixed post, the applied force is now performing two functions. First, it serves to unhook the molecule from the free polymer. (If the free polymer is much larger than the molecule to which the force is applied, then we effectively recover the case of a polymer–post collision.) However, in this scenario, we do no have a fixed contact point; a fraction of the total force F_{contact} is used to generate the center-of-mass motion. We can write the expression for the contact point velocity as (similar to the previous two subsections)

$$v_{\text{contact}} = \frac{F_{\text{contact}}}{\xi_f + \xi_a} \tag{45}$$

where ξ_f and ξ_a are the total friction coefficients of the free polymer and the polymer to which the force is applied (at any given time). This contact-point velocity is defined self-consistently through the solution of the equations of motion for the escape of the two polymers because it depends on the instantaneous lengths of short legs of the molecules, $l_1(t)$ and $l_2(t)$. The force which drives the escape of the chain is due only to the unbalanced length: $f_0(L_1 - l_1(t))/a$. So, for example, in a symmetric hairpin conformation, all of the applied force is contributing to the contact-point velocity. Alternatively, as $l_1(t)$ tends to zero, all of the applied forces tends to contribute to the chain unhooking.

We are currently examining all three systems using molecular dynamics simulations with explicit solvent particles in order to test the predictions from the calculations outlined in this subsection. This work can greatly enhance our fundamental knowledge of the dynamical behavior of polymers in a variety of microenvironments.

V. TETHERED POLYMER CHAINS AND STRONG SHEAR FLOWS

The ideal case of a free, tethered polymer deforming in a strong flow has been studied by Brochard-Wyart and colleagues [43,44]. Although this situation is experimentally relevant (a unique example will be briefly described in the next section), a more common situation is that of a polymer attached to a wall and

subject to a shear flow. Although these two problems appear to be somewhat similar, the dynamics of the polymer attached to the wall is not trivial largely due to the nature of the flow. In the section, we will examine this latter case, which is of particular interest for colloidal stabilization, biological systems [45], and lubrication [46]. We will assume that the wall does not absorb the polymer and that we have a good solvent.

Thus, we will study the case of a polymer attached to a wall and deformed in a shear flow of the type $v_x(y) = \dot{\gamma}y$, where $v_x(y)$ is the flow velocity, $\dot{\gamma}$ is the shear rate, and y is the distance perpendicular to a wall (see Fig. 5). One can generate such a flow by moving one wall at constant speed while keeping the opposite one fixed: The appropriate velocity gradient would then be obtained due to friction between the solvent particles in contact with the plates. A simpler method is to use a pump to create a Poiseuille flow profile inside a tube: If the diameter of the tube is large enough compared to the size of the chain, the velocity gradient near the walls would be essentially linear. The shear flow is characterized by the Weissenberg number $\mathrm{Wi} = \tau\dot{\gamma}$, which is simply the longest relaxation time of the polymer in equilibrium τ times the shear rate $\dot{\gamma}$.

Buguin and Brochard-Wyart [44,47] have studied the extension of the chain near a wall and found that it strongly depends on the value of Wi. These authors described four regimes, shown schematically in Figs. 6b–6e. Very similar regimes were also predicted for a free polymer in a uniform flow [44]. In the present case, however, the polymer is maintained in an asymmetric conformation due to the presence of the wall, and the nature of the flow will lead to subtle fluctuation effects, as we now describe.

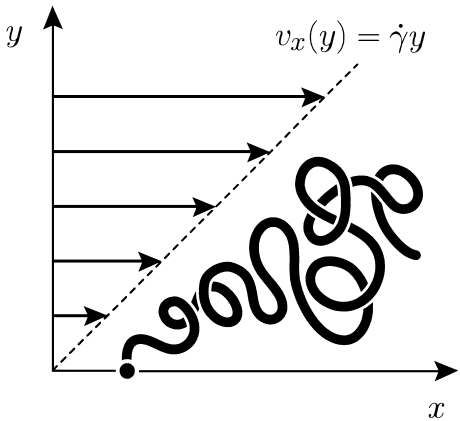


Figure 5 Illustration of a tethered chain in shear flow of type $v_x(y) = \dot{\gamma}y$.

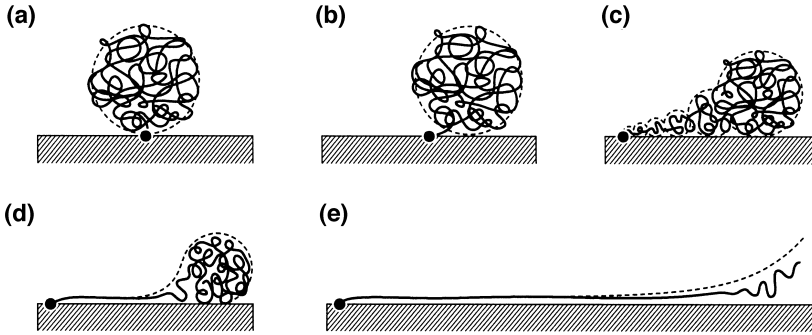


Figure 6 Schematic representation of a polymer chain in the five deformation regimes: (a) When Wi is small, the chain is left unperturbed; (b) at slightly higher velocities, the coil is displaced but not deformed; (c) when Wi increases, the chain first deforms in a series of blobs of increasing size because the tension decreases away from the fixed point (this is called the trumpet conformation); (d) for even larger values of Wi, the tension near the fixed point becomes large enough to fully stretch a part of the chain and we have the so-called stem and flower conformation; (e) finally, large enough shear rates can fully extend the chain and we obtain a rod conformation. (Inspired from Ref. 47.)

A. Fluctuations of a Tethered Polymer in Shear Flow

Doyle et al. [48] (DLV) have reported multiple "fluctuation regimes" for a polymer in the situation shown in Fig. 7. For Wi < 1, the polymer is slightly stretched and undergoes small fluctuations. The second shear flow regime (Wi $\cong$ 1–20) is characterized by an increase in both the mean chain extension and fluctuations about the mean. When Wi > 20, the polymer extends slowly to its maximum extension while fluctuations are reduced compared to the second regime. The interesting fluctuation regime is thus the intermediate one.

Doyle et al. actually found that in the second regime, the fluctuations in the extension of the chain were more than double what is normally observed in equilibrium or in the high-flow regime, a result that was reproduced by simulations [48]. In fact, it was found that the chain was displaying a continuous recirculating motion or a type of cyclic dynamics. As an example, let us take a chain partially elongated near the surface (Fig. 7a). Eventually, it will encounter a particularly strong fluctuation that will bring it away from the wall and into a stronger flow (Fig. 7b). The stronger flow will then lead to an additional extension (Fig. 7c) while the chain is slowly pushed toward the wall by the resulting torque. This rotating motion will take it back into the slower flows found near the wall and the chain will quickly shrink back to a

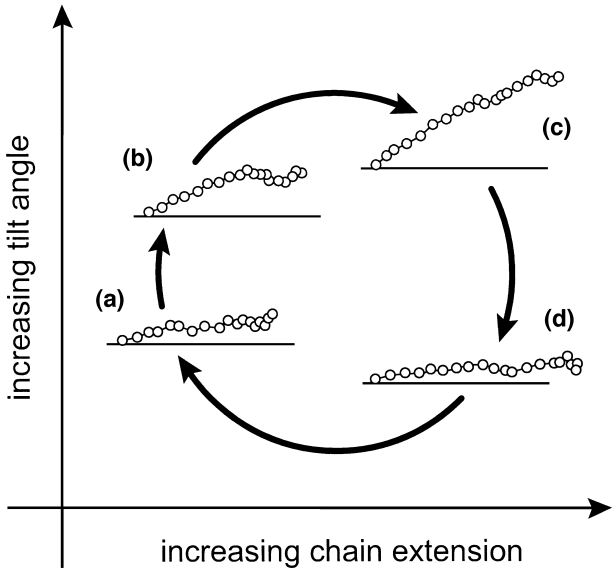


Figure 7 Schematic description of the cyclic chain dynamics found by Doyle et al. [48], as discussed in the text.

less extended state (Fig. 7d) until another strong enough fluctuation restarts the cycle. No specific cycling time has been determined from simulation or experiment. In other words, the cycling time depends entirely on random fluid and polymer fluctuations. This is a remarkably simple and elegant system with nontrivial dynamics.

B. Extensions of a Tethered Polymer in Shear Flow

Ladoux and Doyle [49] have designed models for the FJC and the WLC. In both cases, it is assumed that HIs are negligible (which is not unreasonable because stretched polymers tend to be free draining; furthermore, the presence of the wall partly screens HIs). As a consequence, the friction coefficient of the chain is again assumed to be given by the expression $\xi_R \cong N\xi_1$, where $\xi_1 \cong \eta b$ is the friction coefficient of a monomer of size b.

For the FJC model, these authors used the harmonic approximation [50] to relate the force f and the chain extension h:

$$f = k_{\|}\, h \tag{46}$$

where $k_{\parallel}$ is the longitudinal (entropic) spring constant. Note that we have $k_{\perp} = k_{\parallel}$ in the linear regime. In the transverse direction (y), the thermal motion follows the equipartition law

$$\frac{1}{2}k_{\perp}\delta y^2 = k_B T \tag{47}$$

where δy^2 is the variance of the chain's transverse fluctuations. From Eqs. (46) and (47), we can obtain the coupling between transverse fluctuations and the restoring force:

$$\delta y = \left(\frac{2k_B T h}{f}\right)^{1/2} \sim f^{-1/2} \tag{48}$$

Using Eq. (25), we define the fraction of the total contour length yet to be stretched as

$$\varepsilon \equiv 1 - \frac{h}{h_{\max}} \cong \frac{1}{f} \tag{49}$$

or $f \cong \varepsilon^{-1}$. In the absence of HIs, the equality between the friction and spring forces gives

$$f \cong \xi_R \dot{\gamma} \delta y \tag{50}$$

From Eqs. (48)–(50), we find the scaling relation between the unstretched polymer fraction ε and the shear rate $\dot{\gamma}$ for a FJC:

$$\varepsilon \sim \dot{\gamma}^{-2/3} \tag{51}$$

For the WLC, Eq. (25) leads to the scaling law $f \sim \varepsilon^{-2}$, which leads to

$$\delta y \sim \varepsilon \tag{52}$$

when used with Eq. (48). The WLC scaling law relating ε and $\dot{\gamma}$ is then from Eqs. (48), (50), and (52):

$$\varepsilon \sim \dot{\gamma}^{-1/3} \tag{53}$$

In contrast, Buguin and Brochard-Wyart's result [47] for the stem-and-flower regime predicts

$$\varepsilon \sim \dot{\gamma}^{-1} \tag{54}$$

Ladoux and Doyle [49] have compared their models with Brownian simulations and with experiments using λ-phage DNA molecules. The

simulations agree with the predicted scaling laws in both cases (FJC and WLC). Experimental data also agree with the $\varepsilon \sim \dot{\gamma}^{-1/3}$ scaling law, thus indicating that λ-phage DNA is indeed a WLC. The success of their approach appears to support their assumption that HIs are not important in the high-extension limit. Nevertheless, it is somewhat surprising that taking into account the effect of the flow on the polymer while neglecting the effects of the polymer on the flow (i.e., the model is not self-consistent) provides such a successful model.

VI. POLYMER STRETCHING DURING ELECTROPHORESIS

An interesting example of polymer stretching that has been predicted but is yet to be observed is presented in Figs. 8a, 8c, and 8e. Here, a charged

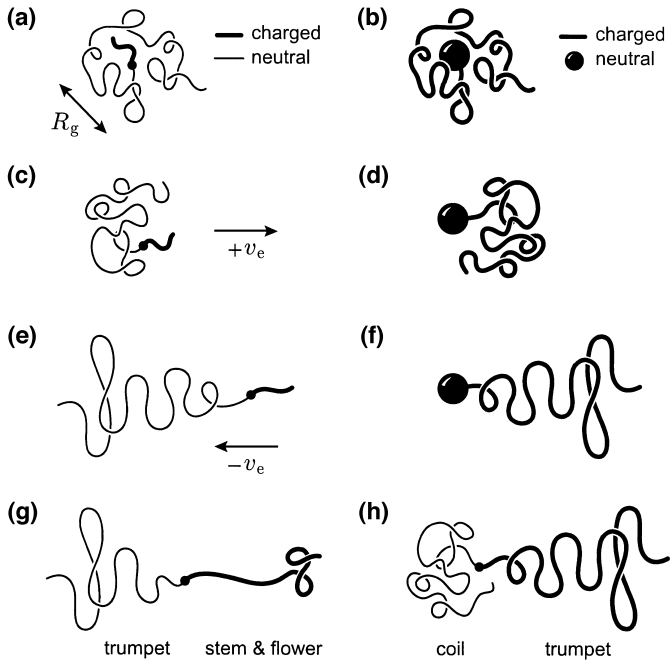


Figure 8 Schematic of free-solution electrophoresis of composite molecules. The hydrodynamic friction forces can deform flexible polymers during the electrophoresis, as discussed in the text.

molecule (e.g., a short DNA fragment) is attached to a long uncharged polymer. If an electric field is applied, the composite molecule will drift under the action of the electrostatic forces, pulling the charged component through the liquid. The neutral polymer simply acts like a "molecular parachute" that slows down the natural electrophoretic drift of the charged molecule. This novel electrophoretic technique, called free-solution conjugate electrophoresis (FSCE), has been successfully used to fractionate neutral polymers [namely polyethylene glycol (PEG)] of varying sizes using small uniform pieces of single-stranded DNA (typically 10–30 monomer long) as the charged "engines" [51,52]. The separation of the various PEG sizes was due to their varying coefficients of friction which resulted in different net electrophoretic velocities.

Because the frictional force generated by the neutral polymer is a function of its conformation, it is important to understand the effect of the electrophoretic motion on the neutral parachute of the composite molecule. As we saw previously, stretched polymers have larger friction coefficients than coiled ones. This is an interesting problem because the neutral polymer stretches in response to the drift (at velocity $+\upsilon_e$) in the immobile fluid: It is as if the fluid is moving at a velocity $-\upsilon_e$ while the polymer is held fixed at its end (see Fig. 8e). On the other hand, the drift velocity would be reduced if the polymer were to stretch. We thus have a self-consistent problem with a negative feedback loop. The only known experimental results were consistent with a model [53,54] where the polymer retained a coil conformation (in other words, the velocity was too low to stretch the "parachute"; see Fig. 8a).

However, as discussed in Ref. 55, one can, in principle, expect a series of conformational deformations similar to the ones predicted by Brochard-Wyart and Buguin [44] and shown in Fig. 6. The first stage would be the *hydrodynamic segregation* of the charged and uncharged molecules (see Fig. 8c). The critical velocity υ_e^* required to achieve this minimal amount of deformation can be estimated using a simple argument. If the charged DNA segment is small compared to the coil of size R_g, segregation is expected to happen when the drag force exceeds the entropic force, keeping the composite molecules as a single object (i.e., in a random coil conformation):

$$\xi \upsilon_c > \frac{k_B T}{R_g} \tag{55}$$

where, for an impermeable coil, $\xi \cong 6\pi\eta R_g$. We thus get $\upsilon_e^* \cong k_B T/6\pi\eta R_g^2$. using typical room-temperature values, we find that this is equivalent to $\upsilon_e^* \cong (2R_{g[\mathrm{nm}]})^{-2}$ m/s, where R_g is in nanometers (nm). We thus need fairly large molecules to have reasonable critical velocities (electrophoretic velocities are generally of the order 0.001–0.1 cm/s). In the stages beyond

segregation (i.e., for even higher velocities) (see Fig. 8e), the trumpet, stem-and-flower, and rod conformations can also appear. As far as we know, this has yet to be observed.

The reverse process consists in attaching uniform uncharged parachutes to a family of charged polyelectrolytes. In this case, electrophoresis is used to fractionate the polyelectrolytes: This method is called free-solution end-labeled electrophoresis, or ELFSE [56]. The relevant theory [54] says that the retardation due to the uncharged molecule is actually like a fluid counterflow. If this is the case, then one also expects molecular deformation (this time of the leading polyelectrolyte engine), as described schematically in Figs. 8b, 8d, and 8f. Again, this rather counterintuitive deformation (pointing against the direction of net motion!) has yet to be observed.

Finally, one may even consider the case of two polymers attached together: one charged and the other not. We expect a global coiled conformation at low velocity. However, as the electric field is increased, we first expect the segregation of the two subcoils. As the field is further increased, we would expect each polymer to go through a series of stretched conformations; this may start occurring for one chain before the other, depending on their properties such as their contour and persistence lengths. Overall, one can imagine no less than 16 different cases, 2 of which are shown in Figs. 8g, and 8h. Again, this situation has yet to be investigated in spite of its potentially rich phase diagram.

VII. CONCLUSION

In this chapter, we have provided an introduction to the topic of polymer stretching and relaxation in good solvents. This is currently a very active area of investigation, both experimentally and theoretically, and hundreds of articles are published every year that treat at least one aspect of this problem. Although our review was certainly not exhaustive (it could not realistically be), it described the main elements typically used to understand the physics of polymer problems when molecular deformations are involved. We also presented four different systems which have attracted much attention over the last several years. These examples are related to the fields of microfluidics and separation science, where polymer entropy, deformation, and relaxation play a central role. In fact, in most of these cases, it is the understanding of these fundamental molecular phenomena that has lead to the design of new devices with improved performances. With the advent of nanofluidics, more work will be needed to understand polymer behavior in the limit where surface interactions dominate and the concept of a "bulk fluid" ceases to make sense.

Some issues remain ill-understood at this point in time. We would like to conclude by discussing two of them. First, current models are usually not self-consistent. For instance, let us take the general problem of a tethered polymer deforming in a given flow (see the previous two sections). This is a classical problem with dozens of variations and applications. In essentially all cases, past theoretical work was aimed at predicting the deformation of the polymer under the specified fluid flow. However, the polymer itself must disturb the flow pattern in its vicinity. Therefore, one should, in principle, treat this aspect of the problem in a *self-consistent* manner. In spite of this possible deficiency, current models appear to reproduce experimental data rather well, which is perhaps a little surprising. This deserves further investigation. Second, current models (except for *molecular dynamics* simulations with explicit solvent) generally assume some kind of effective and instantaneous HI between the monomers (and possibly the walls of the device); for instance, they often use the Oseen tensor to calculate the local effect of HIs [8]. However, HIs cannot propagate at infinite speed. Although it is not clear whether the instantaneous HI approximation can lead to errors for slow processes, one would expect this kind of approach to fail when the phenomenon under study involves short transient effects (e.g., the early stages of polymer stretching that follow the sudden application of a shear flow) or high-frequency periodic perturbations (e.g., a high-frequency oscillatory flow). This is a rather unexplored territory.

REFERENCES

1. Larson, R.G. *The Structure and Rheology of Complex Fluids*; Oxford University Press: New York, 1999.
2. Allen, M.P.; Tildesley, D.J. *Computer Simulation of Liquids*; Oxford Science Publications: New York, 1987.
3. Grest, G.S.; Kremer, K. Phys. Rev. A 1986, *33*, 3628–3631.
4. Rapaport, D.C. *The Art of Molecular Dynamics Simulation*; Cambridge University Press: New York, 1995.
5. Rapaport, D.C. J. Chem. Phys. 1979, *71* (5), 3299–3303.
6. Chu, S. Phil. Trans. R. Soc. Lond. Ser. A: Math Phys. Eng. Sci. 2003, *361*, 689–698.
7. de Gennes, P.G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.
8. Doi, M.; Edwards, S.F. *The Theory of Polymer Dynamics*; Oxford Science Publications: New York, 1986.
9. Teraoka, I. *Polymer Solutions: An Introduction to Physical Properties*; Wiley: New York, 2002.
10. des Cloizeaux, J.; Jannink, G. *Les Polyméres en Solution: Leur Modélisation et leur Structure*. Les Éditions de Physique: Les ULis, 1987.

11. Slater, G.W.; Crisan, S.; Pépin, M.; Gratton, Y. Physics in Canada 2003, *59*, 57–66.
12. Kemp, J.; Chen, Z.Y. Phys. Rev. E. 1997, *56*, 7017–7022.
13. Volkenstein, M.V. *Configurational Statistics of Polymeric Chains*; Interscience: New York, 1963.
14. Slater, G.W.; Hubert, S.J.; Nixon, G.I. Macromol. Theory. Simul. 1994, *3*, 695–704.
15. Glatting, G.; Winkler, R.G.; Reineker, P. Macromolecules 1993, *26*, 6085–6091.
16. Pincus, P. Macromolecules 1976, *9*, 386–388.
17. Marko, J.F.; Siggia, E.D. Macromolecules 1995, *28* (26), 8759–8770.
18. Winkler, R.G. J. Chem. Phys. 2003, *118*, 2919–2928.
19. Han, J.; Turner, S.W.; Craighead, H.G. Phys. Rev. Lett. 1999, *83*, 1688–1691.
20. Han, J.; Craighead, H.G. Science 2000, *288*, 1026–1029.
21. Han, J.; Craighead, H.G. Anal. Chem. 2002, *74*, 394–401.
22. Tessier, F.; Labrie, J.; Slater, G.W. Macromolecules. 2002, *35*, 4791–4800.
23. Tessier, F.; Slater, G.W. Appl. Phys. A 2002, *75*, 285–291.
24. Sebastian, K.L.; Paul, A.K.R. Phys. Rev. E 2000, *62*, 927–939.
25. Sebastian, K.L. Phys. Rev. E 2000, *61*, 3245–3248.
26. Sebastian, K.L. J. Am. Chem. Soc. 2000, *122*, 2972–2973.
27. Sebastian, K.L. Phys. Rev. E 2000, *62*, 1128–1132.
28. Kumar, K.; Sebastian, K.L. Chem. Phys. Lett. 2002, *359*, 101–108.
29. Park, P.J.; Sung, W. J. Chem. Phys. 1999, *111*, 5259–5266.
30. Lee, K.; Sung, W. Phys. Rev. E 2001, *64*, 041801.
31. Lee, K.; Sung, W. Phys. Rev. E 2001, *63*, 021115.
32. Bakajin, O.B.; Duke, T.A.J.; Chou, C.F.; Chan, S.S.; Austin, R.H.; Cox, E.C. Phys. Rev. Lett. 1998, *80* (12), 2737–2740.
33. Doyle, P.S.; Bibette, J.; Bancaud, A.; Viovy, J. Science 2002, *295* (22), 2828–2831.
34. Sevick, E.M.; Williams, D.R.M. Europhys. Lett. 2001, *56* (4), 529–535.
35. Starkweather, M.E.; Muthukumar, M.; Hoagland, D.A. Macromolecules 1998, *31*(16), 5295–5501.
36. Andre, P.; Long, D.; Ajdari, A. Eur. Phys. J B 1998, *4* (1), 307–312.
37. Mayer, P.; Bibette, J.; Viovy, J.L. Mater. Res. Soc. Symp. Proc. 1997, *463*, 529–535.
38. Viovy, J.L. Rev. Mod. Phys. 2000, *72*, 813–872.
39. Batchelor, G.K. J. Fluid. Mech. 1970, *44* (3), 410–440.
40. Barron, A.E.; Harvey, H.W.; Soane, D.S. Electrophoresis 1994, *15*, 597–615.
41. Barron, A.E.; Sonada, W.M.; Blanch, H.W. Electrophoresis 1996, *17*, 744–757.
42. Ekani-Nkodo, A.; Tinland, B. Electrophoresis. 2002, *23* (16), 2755–2765.
43. Brochard, F.; Buguin, A. Comptes. Rendus. Acad. Sci. Ser. II-B 1995, *21*, 463–466.
44. Brochard-Wyart, F.; Buguin, A. *Flexible Polymer Chains in Elongational Flow; Nguyen, T.Q., Kausch, H.H.*; Springer: New York, 1999; 41–66.
45. Lipowski, R. Colloid Surface A 1997, *128*, 255–264.
46. Klein, J.; Kamiyama, Y.; Yoshizawa, H.; Israelachvili, J.N.; Fredrickson, G.H.; Pincus, P.; Fetters, L.J. Macromolecules 1993, *26*, 5552–5560.

47. Buguin, A.; Brochard-Wyart, F. Macromolecules 1996, *29*, 4937–4943.
48. Doyle, P.S.; Ladoux, B.; Viovy, J.L. Phys. Rev. Lett. 2000, *84*, 4769–4772.
49. Ladoux, B.; Doyle, P.S. Europhys. Lett. 2000, *52*, 511–517.
50. Hatfield, J.W.; Quake, S.R. Phys. Rev. Lett. 1999, *82*, 3548–3551.
51. Vreeland, W.N.; Desruisseaux, C.; Karger, A.E.; Drouin, G.; Slater, G.W.; Barron, A.E. Anal. Chem. 2001, *73*, 1795–1803.
52. Vreeland, W.N.; Slater, G.W.; Barron, A.E. Bioconjugate Chem 2002, *13*, 663–670.
53. McCormick, L.C.; Slater, G.W.; Karger, A.E.; Vreeland, W.N.; Barron, A.E.; Desruisseaux, C.; Drouin, G. J. Chromatogr. A 2001, *924*, 43–52.
54. Long, D.; Dobrynin, A.V.; Rubinstein, M.; Ajdari, A. J. Chem. Phys. 1998, *108*, 1234–1244.
55. Desruisseaux, C.; Drouin, G.; Slater, G.W. Macromolecules 2001, *34*, 5280–5286.
56. Ren, H.; Karger, A.E.; Oaks, F.; Menchen, S.; Slater, G.W.; Drouin, G. Electrophoresis 1999, *20*, 2501–2509.

4

Science and Engineering of Nanoparticle–Polymer Composites: Insights from Computer Simulation

FRANCIS W. STARR Department of Physics, Wesleyan University, Middletown, Connecticut, U.S.A.

SHARON C. GLOTZER Departments of Chemical Engineering and Materials Science and Engineering, University of Michigan, Ann Arbor, Michigan, U.S.A.

I. INTRODUCTION

Modification of polymers with an inorganic material to produce a polymer composite is commonplace in the world of modern plastics. Polymer nanocomposites, in which the secondary material has dimensions on nanometer scales, offer exciting opportunities not possible with conventional composites [1–3]. Vast improvements in properties ranging from mechanical response [4] to fire resistance [5,6] can be achieved with small amounts of the nanostructured inorganic additives. As a result, polymer nanocomposites are much lighter than conventional polymer composites while displaying large increases in tensile modulus, strength, toughness and other bulk properties. In addition to their improved properties, polymer nanocomposites are easily molded or extruded, simplifying manufacturing processes. Despite significant progress in materials development, a comprehensive understanding of polymer nanocomposites yielding predictive structure–property relationships is not yet in hand. Without such understanding, progress in nanocomposite development has been largely empirical.

In the last decade, with the explosive development of different kinds of nanoparticles and nanostructured molecules [7–13], such as gold and semiconductor nanoparticles, carbon nanotubes, Buckyballs, polyhedral oligomeric silsesquioxanes, and so forth, many types of polymer nanocomposites have been developed and researched. A common feature in all these hybrid

materials, which distinguishes polymer nanocomposites from traditional composites and filled polymers, is the huge interfacial region resulting from the nanoscopic dimensions of the inorganic component. The enormous amount of interface present means that most of the polymers interact directly with the inorganic component, introducing heterogeneity within the polymer matrix on the nanoscale and deforming the polymer from its bulk conformation. In these unique materials, with only a few volume percent of dispersed nanoparticles, the entire polymer matrix may be considered to be a nanoscopically confined polymer [2].

Polymer nanocomposite performance is defined by three major characteristics [2]: the nanoscale inorganic component, the nanoscopically confined matrix polymer, and the arrangement of the inorganic component within the matrix. Generally speaking, the ultimate properties of a nanoparticle–polymer composite depend as much on the individual properties of the organic and inorganic components as on the relative arrangement and interaction between them. Computer simulations offer a unique opportunity to observe nanoscopic and molecular-level structural and dynamical details in model nanocomposites not readily accessible in experiments. In particular, close examination of the behavior of the polymer chains in the direct vicinity of the nanoparticle, the extent to which the nanoparticle affects the local chain structure and dynamics, and the detailed arrangement of nanoparticles and how it is affected by the strength of polymer–particle interactions, shear, and so forth, can be studied via molecular simulation with coarse-grained interaction potentials. The results of several such studies are reviewed in this chapter [14–16]. We refer the reader to additional computational studies of nanofilled polymers that are not discussed here [17–24].

The chapter is structured as follows. In Section II, we describe the simulation methodology and model we have used to study basic features of nanoparticle-filled homopolymers. In Section III, we describe our findings on the effect of a single-model nanoparticle on the local structure of the polymer chains, and on the local and bulk dynamics. In Section IV, we discuss what our simulations suggest about the role of interactions on nanoparticle dispersion and the role of particle clustering and dispersion on bulk properties like viscosity, and we discuss the possible physical processes controlling nanoparticle clustering. We conclude with a brief discussion of outstanding questions and challenges in Section V.

II. SIMULATION

Specific details of the simulations can be found in Refs. 14–16; we outline the details here for completeness. The first series of simulations [14,15] focused on the structural and dynamic properties of polymer chains in the vicinity of a

single polyhedral nanoparticle. The polyhedral shape of the nanoparticle was chosen for its similarity to existing nanoparticles, such as fullerenes and metallic nanoparticles [25,26]. We also considered this highly symmetric shape because it poses fewer geometrical complications in coding and subsequent analysis than tubelike or platelike nanoparticles (such as carbon nanotubes or organic nanoclays).

Because the goal of these studies was to identify basic physical characteristics of nanoparticle-filled polymer melts that are largely independent of the chemical details of the system, we chose to model the system using very simple potentials that capture only the essential features of the system, such as the chain connectivity, chain statistics, and nanoparticle geometry. Specifically, the polymer chains were represented using a common model in which monomers, interacting via Lennard–Jones interactions, are bonded to nearest-neighbor monomers of the chain via a FENE anharmonic spring potential [27–29]. Similarly, the nanoparticle was represented by a collection of Lennard–Jones particles, bonded in such a way to roughly maintain the icosahedral geometry of the nanoparticle, but allowing for vibrations of the individual force sites around their ideal locations.

The second series of simulations [16] focused on nanoparticle dispersion and its relation to properties. Building upon the previous simulations, the system was modeled using the same potentials, and a range of nanoparticle loadings (volume fractions) and interactions were considered to determine the relation to nanoparticle clustering.

III. EFFECT OF A SINGLE NANOPARTICLE ON POLYMER PROPERTIES

A. Local Structure

To investigate the structural changes in the vicinity of the nanoparticle surface, Refs. 14 and 15 separately examined the behavior of the monomers that make up the chains and the overall behavior of the center of mass of the chains. The monomer structure in the vicinity of the nanoparticle was studied using the monomer density $\rho(d)$ as a function of the distance d from the nanoparticle surface. Because the nanoparticle is nearly spherical, the calculation is simplified by estimating $d = r - r_{\text{surface}}$, where r is the radial position of a monomer relative to the particle center and $r_{\text{surface}} = \frac{1}{12}(42 + 18\sqrt{5})^{1/2}\ell$ is the radius of the inscribed sphere of the icosahedron. We show $\rho(d)$ with (a) attractive and (b) nonattractive (excluded volume only) interactions between monomers and the nanoparticle in Fig. 1.

Due to simple packing constraints, $\rho(d)$ has a well-defined layer structure in both cases. The primary differences in the density profile occur in the vicinity nearest the nanoparticle surface. In the attractive case, the monomer

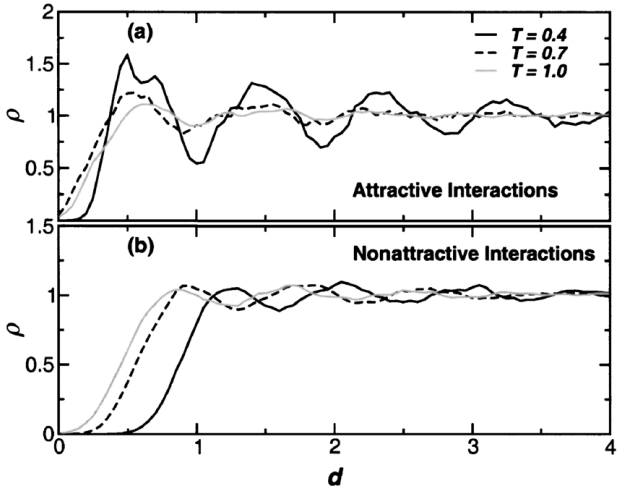


Figure 1 Monomer density profile $\rho(d)$ as a function of the distance from the nanoparticle surface at several temperatures for both (a) attractive (b) nonattractive monomer–particle interactions. (From Refs. 14 and 15.)

density in the first layer is enhanced due to the relatively strong monomer–particle attraction. Additionally, the peak location changes only weakly with temperature, because there is a preferred distance due to the attractions between the nanoparticle and the monomers. The splitting in the first peak at the lowest T studied is due to the fact that the nanoparticle is not spherical, and those monomers near the vertices of the nanoparticle are at a slightly larger distance than those near the center of a face of the nanoparticle. In the case of nonattractive interactions, there is also an enhancement in the density in the first layer, but, in this case, it is attributable to monomers far from the nanoparticle "pressing" the inner monomers toward the interface with the nanoparticle. Moreover, because there is no preferred distance for the innermost monomers, the location of the first peak is more strongly T dependent. The location of the first peak is slightly larger than in the attractive nanoparticle case, and the position of this peak increases with decreasing T. The presence of a maximum and subsequent oscillations in $\rho(d)$ for both the attractive and nonattractive cases is consistent with the density profile of monomers observed near a smooth wall [30–33].

The local packing of the polymer chains can be quantified by examining the radius of gyration:

$$R_g^2 = \frac{1}{M^2}\left\langle \sum_{i,j=0}^{M} (\mathbf{r}_i - \mathbf{r}_j)^2 \right\rangle \tag{1}$$

where M is the number of monomers that make up a chain. To be sensitive to interfacial effects, Ref. 14 calculated R_g as a function of d, the distance of the center-of-mass of a chain from the nanoparticle surface. Furthermore, to gain insight into the orientation of the chains, the radial component $R_g^{\perp}$ of R_g relative to the nanoparticle center was calculated. Insofar as the nanoparticle can be approximated by a sphere, $R_g^{\perp}$ is the component perpendicular to the particle surface. The radial component is found by substituting the segment vector $\mathrm{r}_i - \mathrm{r}_j$ in Eq. (1) with the dot product of the segment vector and the normalized bisector of each chain segment relative to the particle center, yielding

$$R_g^{\perp 2} = \frac{1}{M^2}\left\langle \sum_{i,j=0}^{N} \left(\frac{(\mathbf{r}_i - \mathbf{r}_j)\cdot(\mathbf{r}_i + \mathbf{r}_j)}{|\,\mathbf{r}_i + \mathbf{r}_j\,|} \right)^2 \right\rangle \tag{2}$$

The behavior of R_g^2 and $R_g^{\perp 2}$ is nearly identical for both the attractive and nonattractive monomer–nanoparticle interactions (Fig. 2). The independence of the chain structure on the choice of interactions suggests that changes in

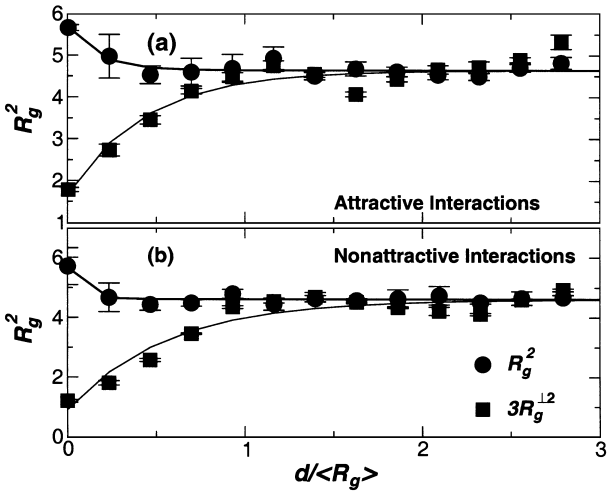


Figure 2 Radius of gyration R_g of the polymer chains as a function of distance $d/\langle R_g \rangle$ of the center of mass of a chain from the nanoparticle surface for $T = 0.4$. The component of R_g perpendicular to the surface is labeled by $R_g^{\perp}$. We show results for both (a) attractive and (b) nonattractive interactions. The increase of R_g, coupled with the decrease of $R_g^{\perp}$, indicates that the chains become increasingly elongated and "flattened" as the surface of the nanoparticle is approached. The effect appears largely independent of the temperature and numerical values of the potential parameters. (From Refs. 14 and 15.)

the structure are primarily due to geometric constraints of packing the chains close to the surface. More specifically, R_g^2 increases by about 25% on approaching the particle surface, whereas $R_g^{\perp 2}$ decreases by slightly more than a factor of 2. These results indicate that the polymers become slightly elongated near the surface and flatten significantly, orienting with the particle surface. At $d \approx R_g$, the chain shape coincides with that of bulk chain. Hence, the surface provides a preferential orientation for the polymers, and it is natural that this effect should persist for a distance R_g, roughly the chain size. We also point out that this result depends only weakly on T. These results are also reminiscent of the properties of polymers near a wall and polymer thin films [30–33], despite the fact that the size of the chains is comparable to that of the nanoparticle, meaning that there is significant curvature in the interface on the scale of the polymer chains.

B. Local and Bulk Dynamics

To elucidate how the polymer dynamics are influenced by the interactions with the surface, it is necessary to consider the melt dynamics as a function of distance from the nanoparticle surface. Single-monomer dynamics can be studied though the incoherent, or self, intermediate scattering, function

$$F_{self}(q,t) = \frac{1}{N}\left\langle \sum_{j=1}^{N} e^{-iq\cdot[\mathrm{r}_j(t)-\mathrm{r}_j(0)]} \right\rangle \tag{3}$$

where q is the wave vector and $\mathrm{r}_j(t)$ is the location of particle j at time t. $F_{\mathrm{self}}(q,t)$ is the Fourier transform of the real-space time-dependent spatial-correlation function. Unfortunately, by this standard definition, $F_{\mathrm{self}}(q,t)$ is averaged over the entire system and hence difficult to relate to surface properties.

To probe the surface properties, Ref. 15 utilized the fact that monomers form well-defined layers surrounding the nanoparticle (see Fig. 1) and split $F_{\mathrm{self}}(q,t)$ into the contribution from monomers located in each layer at $t = 0$. In this way, Eq. (3) can be decomposed as

$$F_{self}(q,t) = \frac{1}{N}\sum_{\mathrm{layers}} N_{\mathrm{layer}}\, F_{\mathrm{self}}^{\mathrm{layer}}(q,t) \tag{4}$$

where N_{layer} is the number of monomers in a given layer.

$F_{\mathrm{self}}^{\mathrm{layer}}(q_0,t)$ is plotted for $T = 0.4$ in Fig. 3. Here, q_0 denotes the value of the first peak in the static structure factor $S(q)$. The relaxation of the monomers closest to the nanoparticle surface is slowest when the monomers and nanoparticle have an attractive interaction. Conversely, when there is no attraction between monomers and the nanoparticle, the relaxation of surface layer

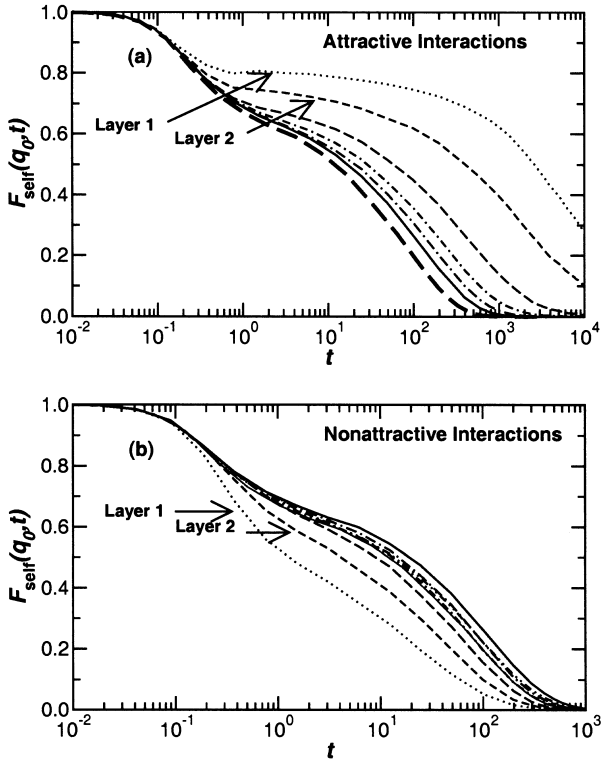


Figure 3 $F_{\text{self}}(q_0,t)$ for the average of all monomers (dotted line) and decomposed into layers (defined by the distance from the nanoparticle surface) for (a) attractive interactions and (b) nonattractive interactions at $T = 0.4$. Layers are defined using the monomer density profile $\rho(d)$ shown in Fig. 1. The minima in $\rho(d)$ define the boundary between layers. In (a), the relaxation near the nanoparticle surface is slowed by roughly two orders of magnitude. In contrast, (b) shows the relaxation of $F_{\text{self}}(q_0,t)$ is enhanced by roughly one order of magnitude near the surface. The relaxation time of the outermost layer in both cases nearly coincides with the relaxation time of the pure system. (From Refs. 14 and 15.)

monomers is significantly enhanced in comparison to the bulk. The altered dynamics persist for a distance slightly less than $2R_g$ (radius of gyration) from the surface. The systems shows a two-step relaxation at this T because the relaxation occurs through a fast vibrational motion and a slower structural rearrangement that occurs on a significantly longer timescale. Figure 3 also shows the relaxation of $F_{\text{self}}(q_0, t)$ for a pure system at the same density and temperature. The results for the pure system nearly coincide with the behavior

of the outermost layer, regardless of the choice of monomer–particle interactions, indicating that, at large distances, bulklike behavior is recovered. The impact of the surface roughness on the surface dynamics has also recently been considered [34].

These changes in the local dynamics naturally also have an effect on the overall bulk dynamics of the system. Far from the nanoparticle surface, the dynamics of these systems nearly coincides with those of a pure system, and so the changes in bulk properties depend only on the change in the surface properties. Hence, the system with attractive nanoparticle-polymer interactions is expected to have a longer average relaxation time, and the system without nanoparticle-polymer attraction should have a shorter average relaxation time. This can be clearly illustrated by the relaxation time τ of the spherically averaged intermediate scattering function

$$F(q,t) \equiv \frac{1}{NS(q)} \left\langle \sum_{j,k=1}^{N} e^{-iq\cdot[\mathrm{r}_k(t)-\mathrm{r}_j(0)]} \right\rangle \tag{5}$$

$F(q,t)$ is normalized by the structure factor $S(q)$ so that $F(q,0) = 1$. The relaxation time τ is defined by $F(q,\tau) \equiv 0.2$. The exact choice for the definition of τ does not qualitatively affect the results. At each T, τ is larger than that of the pure system for the attractive system, and this difference

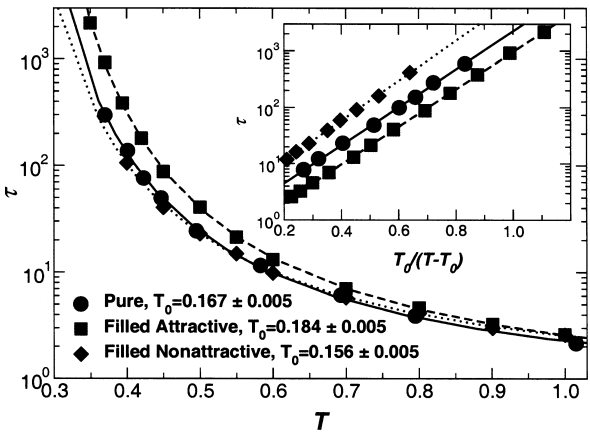


Figure 4 Temperature dependence of the relaxation time of the intermediate scattering function. The lines are a fit to the Vogel–Fulcher–Tammann (VFT) form of Eq. (6). The inset shows the same data plotted against reduced temperature $T_0/(T - T_0)$ to show the quality of the VFT fit. For clarity in the inset, τ of the pure system is multiplied by 2, and τ of the filled nonattractive system is multiplied by 4. (From Refs. 14 and 15.)

increases with decreasing T (Fig. 4). Conversely, τ in the nonattractive system is slightly smaller at low T than in the pure system, and this difference increases with decreasing T. These observations can be related to the bulk properties of the system through the glass transition temperature T_g. Given the relative changes in τ, the attractive system is expected to have greater T_g than the pure system, and the nonattractive system should have a T_g less the pure system.

References 14 and 15 estimated changes in T_g by fitting τ to the Vogel–Fulcher–Tammann (VFT) form

$$\tau \sim e^{A/(T-T_0)} \tag{6}$$

where T_0 is typically quite close to the experimental T_g value [35], provided the data fit is at sufficiently low T; hence, changes in T_g are reflected in T_0. T_0 increases in the system with attractive nanoparticle-polymer interactions, but decreases in the system with only an excluded volume interaction. Thus, the effect of the steric hindrance introduced by the nanoparticle decreases $\tau(T)$ and T_0, in spite of the fact that monomers have a reduced number of directions in which to move and, hence, reducing the degrees of freedom that aid in the loss of correlations. The fact that T_0 shifts in opposite directions for attractive versus purely excluded volume interactions further demonstrates the importance of the surface interactions. Similar results for the bulk viscosity have also been found [21]. Other systems, like thin films [33] and brushes [36], show similar dynamical phenomena.

IV. COMPOSITE MATERIALS

A. Role of Interactions on Particle Clustering

Nanoparticle clustering can be controlled through a variety of parameters, such as concentration and temperature. One of the most fundamental parameters affecting the tendency of nanoparticles to aggregate or disperse is the interactions between the nanoparticles and the interactions between nanoparticles and the surrounding polymer matrix. Reference 16 determined the range of interactions for which clustering or dispersion occurs in a simple model system and, in doing so, provides some insight into the mechanism controlling clustering. In the following, we summarize how varying polymer–nanoparticle interactions while keeping the loading volume fraction of nanoparticles, ϕ, fixed alters the clustering properties of the nanocomposite. Additionally, varying the monomer–particle interaction strength $\varepsilon_{\mathrm{mp}}$ and holding T fixed eliminates changes in clustering that arise from simple vibrational excitation. Such vibrational changes can obscure changes in the properties due to changes in the state of particle clustering.

In order to track changes in nanoparticle clustering, a simple and reliable metric indicating the qualitative changes in clustering is needed. As particles cluster, there are significant changes in the particle–particle and particle–monomer components of the potential energy, reflecting the difference in the number of particle–particle or particle–monomer contacts in the system. Hence, Ref. 16 used the potential energy per force site between nanoparticles u_{pp} as an indicator of nanoparticle clustering. Additionally, the specific heat $c_V^{pp} = \partial u_{pp}/\partial T$ aids in the identification of the crossover point between the asymptotic limits of clustering and dispersion.

Figure 5a shows u_{pp} as a function of the monomer–particle interaction strength ε_{mp} at fixed $\phi = 0.094$ and $T = 2.0$. At large and small ε_{mp}, u_{pp} is nearly constant, but it makes a relatively abrupt crossover between two extremes over a narrow range $1.25 < \varepsilon_{mp} < 1.35$. Visual inspection of the configurations (Fig. 6) at the extreme values of ε_{mp} confirms that the low values of u_{pp} correspond to clustered states, whereas the larger values correspond to dispersed configurations. Hence, for this (T,ϕ), ε_{mp} needs to be only slightly stronger than $\varepsilon \equiv 1$ (the strength of monomer–monomer interactions) for the particles to disperse. The range over which the crossover occurs is more

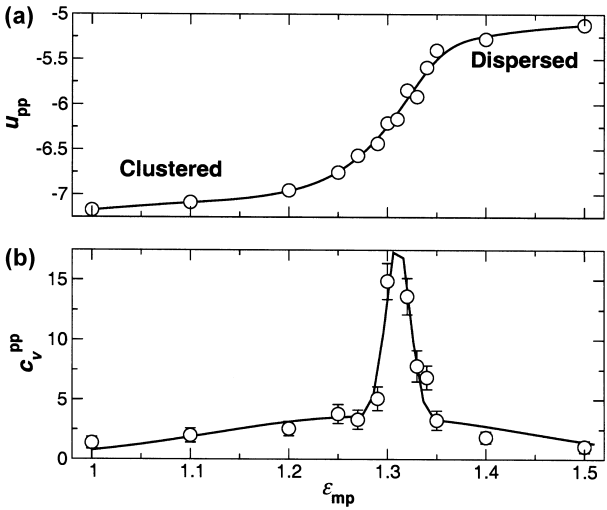


Figure 5 (a) The nanoparticle component of the potential energy u_{pp} shows the crossover between clustered (low u_{pp}) and dispersed (high u_{pp}) states. Because the T is held constant, changes in u_{pp} are dominated by the clustering of the nanoparticles. (b) The specific heat c_V^{pp} maximum provides a more reliable estimate of the crossover point between clustered and dispersed states. (From Ref. 16.)

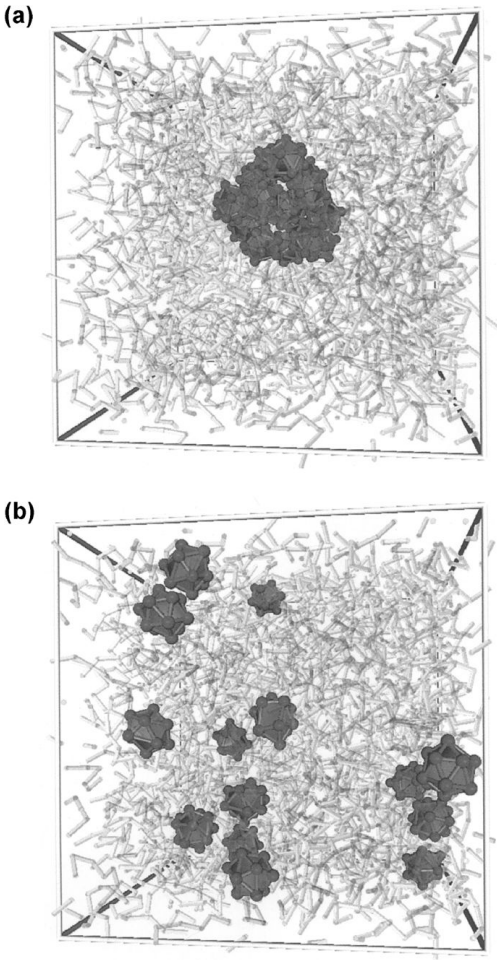


Figure 6 Snapshots of the simulated nanocomposite for $T = 2.0$ and $\phi = 0.046$. Part (a) shows that the nanoparticles are clustered for a relatively weak interaction $\varepsilon_{mp} = 1.0$. Similarly, (b) shows dispersion occurs for $\varepsilon_{mp} = 1.5$. The chains are transparent to avoid obscuring the nanoparticles. The small spheres on the nanoparticles represent the locations of the Lennard–Jones force sites. (From Ref. 16.)

clearly shown by c_V^{pp} (Fig. 5b). Moreover, very small changes in the relative interactions between nanoparticles and polymer can cause dramatic qualitative changes in the structure of the system, which, in turn, must be reflected in the bulk properties of the material.

B. Relating Material Structure and Properties

Development of new materials can be accelerated by understanding how the ultimate material properties are connected to the material structure. Thus, it is desirable to understand the way in which a rheological property like viscosity η varies as ε_{mp} changes and hence reflects the degree of clustering. Viscosity is given by

$$\eta(\dot{\gamma}) = \frac{\langle P_{xy} \rangle}{\dot{\gamma}} \tag{7}$$

where $\langle P_{xy} \rangle$ is the average of the component of the pressure tensor along the flow and gradient directions of the shear, and γ is the shear rate [16]. Reference 16 evaluated η at fixed $\phi = 0.172$ and $\dot{\gamma} = 0.01$ — a low enough shear rate that the system is Newtonian [16]. Figure 7a shows that η appears to approach nearly constant values at $\varepsilon_{mp} = 1$ and 1.5, with a gradual crossover around $\varepsilon_{mp} \approx 1.3$. In addition, c_V^{pp} of the sheared systems, shown in Fig. 7b, indicates the crossover in clustering behavior occurs in the same range of ε_{mp} that η changes between asymptotic regimes. Thus, η is more sensitive to the state of particle clustering than to variation of ε_{mp}.

To take advantage of the fact that increased dispersion leads to an increase in viscosity, we also need to have some understanding of why this should be the case. In fact, hydrodynamic arguments would predict the opposite effect — that a large or extended rigid body embedded in a fluid results in a greater viscosity than a disperse collection of small rigid bodies [37]. However, we must also consider the changes in the dynamics of the polymers near the surface of the nanoparticles, as discussed in the previous section. For attractive interactions, such as we have here, we know that relaxation times of surface monomers increase, corresponding to increase of the "local viscosity." For a well-dispersed configuration, the amount of exposed nanoparticle surface grows linearly with the number of nanoparticles; if clustering occurs, the amount of exposed surface grows sublinearly relative to the number of particles. Because we expect that the changes in η are proportional to the amount of polymers that are in contact with the nanoparticles, η should be larger for well-dispersed configurations than for clustered configurations.

A test of the expected correlation with the exposed surface area A is estimated by the fraction of nanoparticle force sites in contact with a chain (Fig. 7c). There is a large change in A in the same range as the large changes

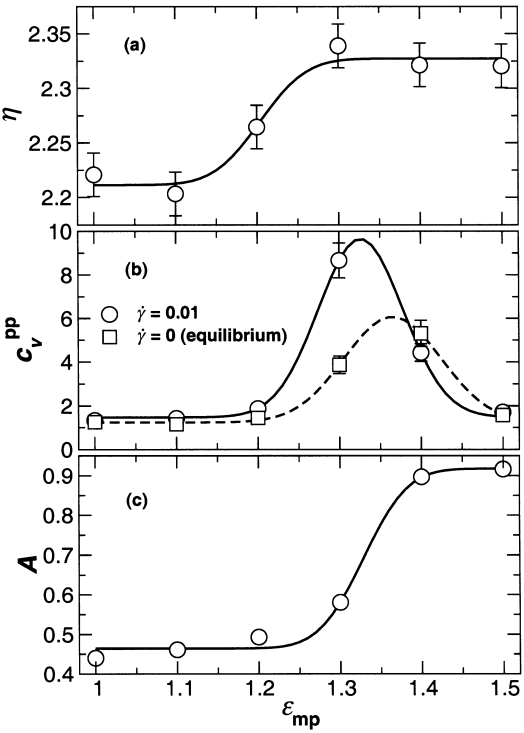


Figure 7 (a) Viscosity η as a function of ε_{mp} at fixed $T = 2.0$ and $\phi = 0.172$. (b) Specific heat calculated for the sheared configurations (circles, solid line) and equilibrium configurations (squares, broken line). Note the shift in the maximum and its location compared with the crossover in the behavior of η. (c) The fraction of nanoparticle-force sites in contact with a polymer, an estimate of the exposed surface area A. The lines serve only as guides to the eye. (From Ref. 16.)

in η. However, we note that that the regions where A and η significantly change do not match precisely. This may be an indication that less pronounced hydrodynamic effects do play a role. Some progress in accounting for hydrodynamic effects on viscosity in a system with a single nanoparticle has been made [21], but these results do not account for changes due to clustering in nondilute nanoparticle composites.

C. Physical Mechanism Controlling Clustering

In this subsection, we discuss the possible physical process though which particle clustering occurs. There are a variety of possible mechanisms, the

most obvious of which is that particles cluster via ordinary phase separation, as in a binary mixture. Phase separation of this type is first order, and so it is expected that u_{pp} and c_V^{pp} exhibit a discontinuity at the transition line, provided we do not follow a path through the critical point. We can see that, in Fig. 5, the crossover from dispersed to clustered states appears to occur through a continuous process. At first glance, this would argue against phase separation, but we must consider that, because of the finite size of the system, we expect rounding of the transition [38]. Therefore, while these results are not inconsistent with phase separation, we need further evidence to make a compelling case for, or against, phase separation.

To obtain such evidence, the system was cooled from the dispersed to the clustered phase for several different ϕ at fixed $\varepsilon_{mp} = 1.3$. In doing so, Ref. 16 estimated the approximate phase boundary, thereby determining if the behavior of the specific heat matches the expectations for phase separation, namely that c_V^{pp} should be increasing toward divergence as we approach the critical point of the transition. In Fig. 8, we show the behavior of u_{pp} and c_V^{pp} as a function of T for several loadings. The peak in c_V^{pp} facilitates identification of a clustering boundary. The temperature T^* of the boundary between clustered and dispersed states is approximated by the location of the

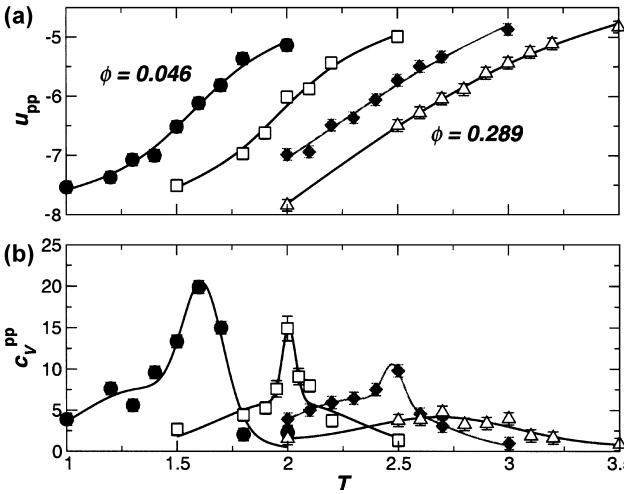


Figure 8 (a) Potential energy u_{pp} and (b) specific heat c_v^{pp} as a function of T for loadings $\phi = 0.046$ (○), 0.094 (□), 0.172 (◇), and 0.289 (△). Nanoparticles are clustered for low u_{pp}, and the approximate boundary between clustered and dispersed states is given by the maximum in c_v^{pp}. The lines are only a guide to the eye. (From Ref. 16.)

maximum in c_V^{pp}. Figure 9 shows that the boundary is positively sloped, indicating that clustering occurs for large ϕ and low T. If the clustering mechanism were analogous to binary phase separation, the critical point must be at some $\phi > 0.3$, because the boundary is positively sloped to that ϕ. Hence, the maximum value of c_V^{pp} should be increasing as ϕ increases toward possible critical point. However, Fig. 8b shows that the amplitude of the peak in c_V^{pp} decreases and becomes broader as we increase ϕ. Hence, the behavior of c_V^{pp} appears to be contrary to our expectations for phase separation.

If nanoparticle clustering is not governed by a phase-separation process, then what is the controlling mechanism? The decrease in the amplitude of the c_V^{pp} peak with increasing ϕ is consistent with the predicted behavior for an associating system [39,40]. The model of equilibrium polymerization [40] also predicts that the loci of specific heat maxima should shift location according to

$$\phi \sim \exp\left(-\frac{E_1}{T^*}\right) \tag{8}$$

Within the limits of uncertainty in the data, the clustering boundary shown in Fig. 9 can be described by Eq. (8). These findings suggest that the clustering transition in this system, and presumably in many similar real nanocomposite systems characterized by short-range, van der Waals-type interactions, is

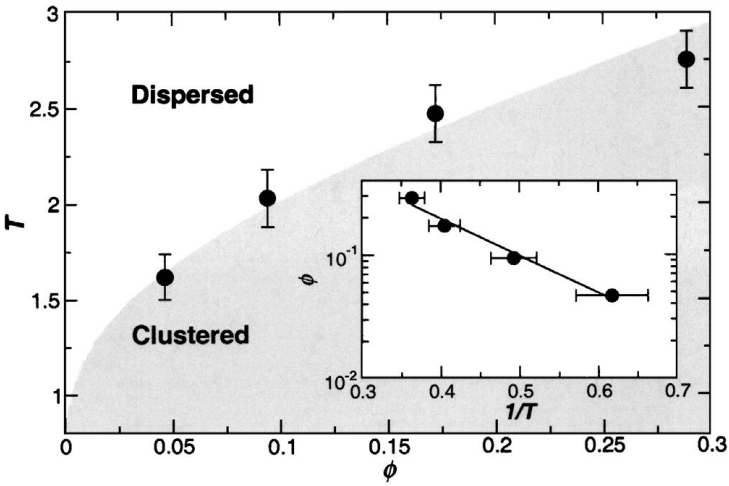


Figure 9 The "clustering diagram" of the nanoparticles, as a function of T and ϕ. The boundary of the shaded region is determined by fitting the points using Eq. (8), shown in the inset. (From Ref. 16.)

controlled by the same mechanism as simple associating systems for the values of ϕ studied here. This observation provides a framework for rationalizing the behavior of many nanoparticle systems, which should, in turn, aid in the control of dispersion and nanocomposite properties.

V. CONCLUSION

In this chapter, we have reviewed the results of molecular dynamics simulations of a coarse-grained model of a nanoparticle-filled polymer melt. These simulations provide insight into how polymer–nanoparticle interactions can affect structural, dynamical, thermodynamic, and rheological properties of the matrix polymer. The scientific understanding obtained by our results can be used to guide the engineering of polymer nanocomposites through the manipulation of the interfacial interactions between the organic and inorganic components. For more detailed insight on a specific polymer–nanoparticle composite, atomistic simulations with more realistic force fields are necessary. However, such simulations alone cannot be performed for large systems due to the computational cost of including such a high level of detail. Thus multiscale approaches are needed, whereby electronic and atomistic detail obtained from a combination of ab initio and atomistic simulation techniques are fed upward to parameterize models for coarse-grained simulations such as those described here [41]. In turn, output from the coarse-grained simulations can be fed into, for example, field-theoretic simulations for the prediction of microstructures resulting from thermodynamic immiscibility of nanoparticle-filled polymers and blends [42–44]. These microstructures can be fed into macroscopic models for calculation of bulk properties like conductivity and modulus [45].

There is increasing evidence that nanoparticle shape can play a large role in dictating properties of the matrix in which the nanoparticles are dispersed. Clay platelets, for example, can induce different degrees of changes than spherically symmetric nanoparticles composed of the same material [46–48]. Recent simulations of tethered nanoparticles, in which nanoparticles are not dispersed freely in a matrix but rather are combined with polymeric tethers through covalent bonding of the tethers to strategic locations on the nanoparticle surface, indeed demonstrate the importance of nanoparticle shape in controlling local nanoparticle packing when nanoparticles and tethers are immiscible [49]. Functionalization of nanoparticles and inorganic nanostructured molecules by cross-linking polymers is another approach to engineering polymer–nanoparticle nanocomposites [50,51]. Here, too, simulations can provide insight via a systematic investigation of the local structure and mechanical properties and how they depend on such characteristics as polymer molecular weight, stiffness, and functionality of the nanoparticle [52].

REFERENCES

1. MRS Bulletin, May 2001 issue, Volume 26.
2. Vaia, R.A.; Giannelis, E.P. MRS Bull. 2001, *26*, 394.
3. Vaia, R.A.; Giannelis, E.P. Eds. *Polymer Nanocomposites*; American Chemical Society: Washington, DC, 2001.
4. Giannelis, E.P. Adv. Mater. 1996, *8,* 29.
5. Kashiwagi, T.; Grulke, E.; Hilding, J.; Harris, R.; Awad, W.; Douglas, J. Macromolecules 2002, *23*, 761.
6. Tang, Y., et al. Polym. Int. 2003, *52*, 1396.
7. Murray, C.B.; Kagan, C.R.; Bawendi, M.G. Annu. Rev. Mater. Sci. 2000, *30*, 545.
8. Busbee, B.D.; Obare, S.O.; Murphy, C.J. Adv. Mater. 2003, *15*, 414.
9. Sun, Y.G.; Xia, Y.N. Science 2002, *298*, 2176.
10. Pinna, N.; Weiss, K.; Urban, J.; Pileni, M-P. Adv. Mater. 2001, *13*, 261.
11. Kooij, F.M.V.; Kassapidou, K.; Lekkerkerker, H.N.W. Nature 2000, *406*, 868.
12. Manna, L.; Milliron, D.J.; Meisel, A.; Scher, E.C.; Alivisatos, P.A. Nature Mater. 2003, *2*, 382.
13. Jin, R.; Cao, Y.; Mirkin, C.A.; Kelly, K.L.; Schatz, G.C.; Zheng, J.G. Science 2001, *294*, 1901.
14. Starr, F.W.; Schrøder, T.B.; Glotzer, S.C. Phys. Rev. E 2001, *64*, 21,802.
15. Starr, F.W.; Schrøder, T.B.; Glotzer, S.C. Macromolecules 2002, *35*, 4481.
16. Starr, F.W.; Douglas, J.F.; Glotzer, S.C. J. Chem. Phys. 2003, *119*, 1777.
17. Vacatello, M. Macromolecules 2001, *34*, 1946.
18. Vacatello, M. Macromolecules 2002, *35*, 8191.
19. Vacatello, M. Macromol. Theor. Simul. 2002, *11*, 757.
20. Salaniwal, S.; Kumar, S.K.; Douglas, J.F. Phys. Rev. Lett. 2002, *89*, 258301.
21. Smith, G.D.; Bedrov, D.; Li, L.; Byutner, O. J. Chem. Phys. 2002, *117*, 9478.
22. Smith, J.S.; Bedrov, D.; Smith, G.D. Compos. Sci. Technol. 2003, *63*, 1599.
23. Brown, D.; Mélé, P.; Marceau, S.; Albérola, N.D. Macromolecules 2003, *36*, 1395.
24. Gersappe, D. Phys. Rev. Lett. 2002, *89*, 058301.
25. Cole, D.H.; Shull, K.R.; Baldo, P.; Rehn, L. Macromolecules 2000, *32*, 771; Carusoa, F.; Mohwald, H. Langmuir 1999, *15*, 8276; Sarathy, K.V.; Narayan, K.S.; Kim, J.; White, J.O. Chem. Phys. Lett. 1999, *318*, 543.
26. Barnes, K.A.; Karim, A.; Douglas, J.F.; Nakatani, A.I.; Gruell, H.; Amis, E.J. Macromolecules 2000, *33*, 4177.
27. Bird, R.B.; Curtiss, C.F.; Armstrong, R.C.; Hassager, O. *Dynamics of Polymeric Liquids: Kinetic Theory*; Wiley: New York, 1987; Vol. 2.
28. Grest, G.S.; Kreme, K. Phys. Rev. A 1986, *33,* 3628.
29. Rudisill, J.W.; Cummings, P.T. Rheol. Acta 1991, *30*, 33.
30. Kumar, S.K.; Vacatello, M.; Yoon, D.Y. J. Chem. Phys. 1989, *89*, 5206.
31. Kumar, S.K.; Vacatello, M.; Yoon, D.Y. Macromolecules 1990, *23*, 2189.
32. Wang, J-S.; Binder, K. J. Phys. I (France) 1991, *1*, 1583.
33. Mischler, C.; Baschnagel, J.; Binder, K. Adv. Colloid Interf. Sci. 2001, *94*, 197. *and references therein*.

34. Smith, G.D.; Bedrov, D.; Borodin, O. Phys. Rev. Lett. 2003, *90*, 226, 103.
35. Debenedetti, P.G. *Metastable Liquids*; Princeton University Press: Princeton, NJ, 1996.
36. Savin, D.A.; Pyun, J.; Patterson, G.D.; Kowalewski, T.; Matyjaszewski, K. J. Polym. Sci. Polym. Phys. 2002, *40*, 2667.
37. Bicerano, J.; Douglas, J.F.; Brune, D.A. Rev. Macromol. Chem. Phys. 1999, *C39*, 561.
38. Binder, K. Rep. Prog. Phys. 1987, *50*, 783.
39. Green, S.C. Annu. Rev. Phys. Chem. 2002, *53*, 173.
40. Dudowicz, J.; Freed, K.F.; Douglas, J.F. J. Chem. Phys. 1999, *111*, 7116.
41. Glotzer, S.C.; Paul, W. Annu. Rev. Mater. Res. 2002, *32*, 401.
42. Lee, B.P.; Douglas, J.F.; Glotzer, S.C. Phys. Rev. E 1999, *60*, 5812.
43. Ginzburg, V.V.; Balazs, A.C. Macromolecules 1999, *32*, 5681.
44. Ginzburg, V.V.; Singh, C.; Balazs, A.C. Macromolecules 2000, *33*, 1089.
45. Starr, F.W.; Glotzer, S.C. Materials Research Society Fall '00 Proceedings: Filled and Nanocomposite Polymer Materials, 2001; 661.
46. Schmidt, G.; Malwitz, M.M. Curr. Opin. Colloid Interf. Sci. 2003, *8*, 103.
47. Luo, J.J.; Daniel, I.M. Compos. Sci. Technol. 2003, *63*, 1607.
48. Fornes, T.D.; Paul, D.R. Polymer 2003, *44*, 4993.
49. Zhang, Z.L.; Horsch, M.A.; Lamm, M.H.; Glotzer, S.C. Nano Letters 2003, *3,* 1341.
50. Soles, C.L.; Lin, E.K.; Wu, W.L.; Zhang, C.; Laine, R.M. *Organic/Inorganic Hybrid Materials*; Materials Research Society: Warrendale, Pa, 2000; Vol. 628.
51. Choi, J.; Yee, A.F.; Laine, R.M. Macromolecules 2003, *36*, 5666.
52. Lamm, M.H.; Chen, T.; Glotzer, S.C. Nano Letters 2003, *3* (8), 989.

5

Polymeric Additives as Modifiers of Hydrocarbon Crystallization Behavior

JEFFREY L. HUTTER The University of Western Ontario, London, Ontario, Canada

I. INTRODUCTION

Crude oils and most of their commercial derivatives such as diesel fuels and fuel oils are complex fluids containing normal and branched alkane molecules, aromatic compounds, and other polar and nonpolar compounds. Of these, the long-chain normal alkane molecules, C_nH_{2n+2}, (also known as normal paraffins or simply as waxes) play an important role in controlling the rheological properties of these mixtures, particularly at low temperatures. As the temperature of such a solution is lowered, the larger *n*-alkanes (those with $n \geq 20$) precipitate from solution, forming a weakly linked, gellike network of crystals, drastically increasing its viscosity. This causes difficulties in the production, transport, and storage of crude oils and fuels, particularly in cold climates. An important case where such concerns arise is the use of diesel fuels in winter.

This chapter will explore the techniques that have been employed to control the crystallization of *n*-alkanes from crude oils and fuels. In Section II, a background for the particular problem of *n*-alkane crystallization from petroleum fuel is provided. Crystallization theory, particularly for the *n*-alkanes, will be briefly reviewed in Section III, whereas the effects of additives on crystallization will be discussed in Section IV. With this background, Section V describes the experimental and theoretical approaches that have been used to study the particular problem of *n*-alkane crystallization from complex fluids. Finally, Section VI will conclude with some suggestions for work that remains to be done.

II. BACKGROUND

It is well known that the larger *n*-alkane molecules begin to precipitate from crude oils, fuel oils, and diesel fuels as temperatures drop below room temperature. This results in viscous properties that are strongly sensitive to temperature. In order to characterize the behavior of crude oils and their derivatives, simple figures of merit have been developed: the *cloud point* (CP), when crystals first start to form, increasing optical scattering from the solution [1]; the *cold-filter plug point* (CFPP), below which the fluid will no longer pass through a 45-μm grid due to blockage by small crystals [2]; and the *pour point* (PP), below which the fluid will no longer flow under gravity [3]. Usually, these temperatures occur in the order $T_{CP} > T_{CFPP} > T_{PP}$. Requirements for diesel fuel operability are typically imposed in terms of the CFPP or, in North America, in terms of the *low-temperature flow test* (LTFT) [4]. Although such phenomenological values are often sufficient to characterize those properties of a crude oil or fuel that are relevant to a specific application, such as transport of crude oils through pipelines or delivery of fuels to combustion chambers, they lack the predictive power required to develop procedures for improving these properties.

Microscopic studies of cooled crude and fuel oils have demonstrated that it is the formation of a network of *n*-alkane crystals that is responsible for the sharp rise in viscosity at low temperatures [5]. Thus, in order to improve the performance of diesel fuels and fuel oils at low temperature, an understanding of the precipitation of *n*-alkanes is needed. Diesel fuels, for instance, typically contain of the order of 10–30% *n*-alkane molecules by mass, with sizes up to at least $n = 30$ and a mass average n of 15–18 [6–8]. With depletion of high-quality reserves and increased demand for the lighter *n*-alkanes to produce jet fuels, the heavy *n*-alkane content of diesel fuels is likely to increase with time [9].

Figure 1 shows a typical *n*-alkane distribution in a diesel fuel. As the temperature is lowered, the heavier *n*-alkane molecules (which are the least soluble molecules present) can precipitate from the fuel, typically forming 2 wt% solids at a temperature of −10°C [9,10]. Also shown are the *n*-alkane distributions in both the crystallized fraction and the residual fluid for the same fuel at −10°C. Because the *n*-alkanes tend to crystallize into thin, platelike crystals, as shown in Fig. 2, this precipitate is very effective at blocking filters and narrow passages, such as those found in fuel injectors. This leads to the eventual blockage of the fuel lines and filters, preventing flow of diesel into engine combustion chambers and halting operation of the engine.

To alleviate the problem of *n*-alkane precipitation from diesel fuels, the formation of crystals large enough to plug narrow passages must be pre-

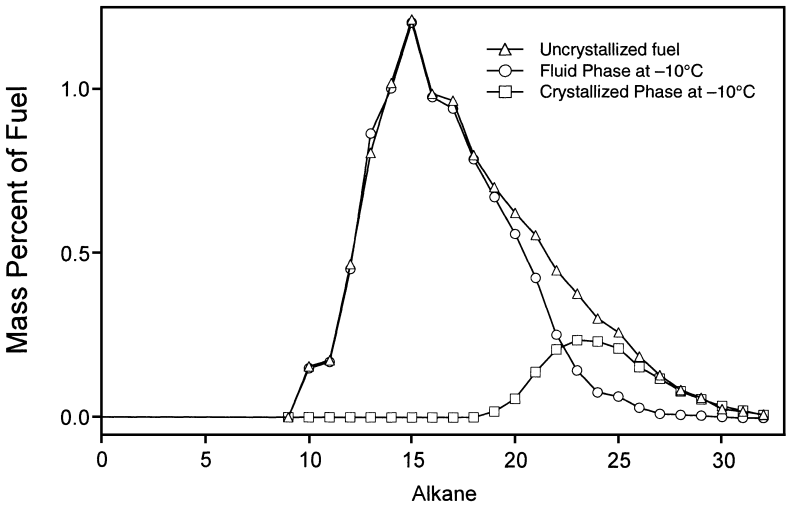


Figure 1 Distribution of *n*-alkanes in diesel fuels: triangles, distribution by alkane length in uncrystallized fuel; circles, distribution in fluid phase at −10°C; squares, distribution in crystallized phase at −10°C. (Data adapted from Ref. 10.)

vented. One traditional approach has been to blend the fuel with lighter fluids, such as kerosene [9,11]: typically, a 2°C decrease in the CP [12] and ~1°C decrease in the CFPP and PP [5,7,11] can be achieved per 10% (by volume) dilution. Such thermodynamic approaches have the disadvantage of expense (such diluents are often more expensive than the base fuel) and possible changes in engine operating characteristics [7]. Another approach is to heat the fuel via engine heat, using return lines. However, this is only effective during engine operation. It has also been suggested that the fuel tank itself be heated [13], although this has not been a widely adopted approach.

Rather than such thermodynamic approaches, attention recently has focused on the use of additives [14], which are termed *flow improvers* and *pour-point depressants* for their operational effects. These additives are effective at very low concentrations—a few hundred parts per million can yield effects on flow properties equivalent to dilution by 10% kerosene [11]—and are relatively inexpensive. Such additives appear to change the *kinetics* but not the *thermodynamics* (i.e., melting point or equilibrium crystal structure) of crystallization and so have become collectively known as kinetic inhibitors. Kinetic inhibitors are usually divided into two classes: *nucleation promoters*, which increase the rate of nucleation of wax crystals, and *crys-*

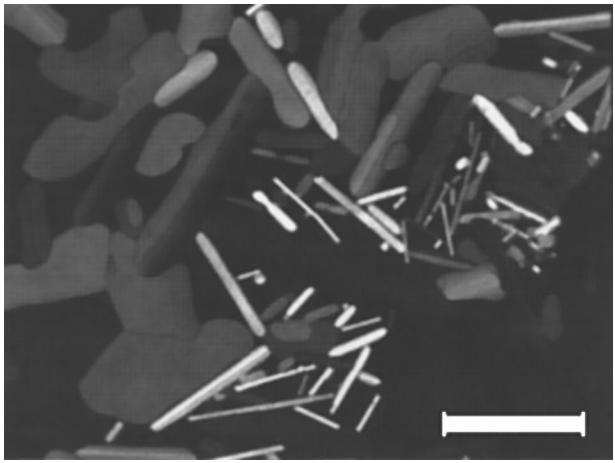
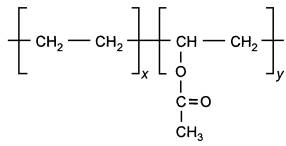


Figure 2 Polarizing light microscope image of $C_{23}H_{48}$ crystals grown from a 20% solution with $C_{12}H_{26}$. Note the nearly edge-on crystals, which illustrate the high aspect ratio (at least 25 : 1 in this image) of these crystals. The scale bar (lower right) indicates 100 μm.

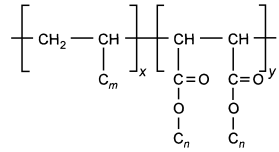
tallization inhibitors, which slow the growth of existing crystals, although there is overlap between the classes. Somewhat paradoxically, nucleation promoters and crystallization inhibitors are often used simultaneously. This results in the formation of many small crystals, rather than a few large ones, improving passage of the crystals through fuel lines and filters.

Most of the kinetic inhibitors in use are polymeric compounds. Crystallization inhibitors discussed in the literature are often copolymers based on poly(vinyl acetate), such as random ethylene–vinyl acetate copolymers (EVAs) [7,15,16], which also seem to serve as nucleation promoters, and alternating fumarate–vinyl acetate copolymers (FVAs) [16–19]. Other examples include poly(alkylacrylates), such as poly(octadecyl methacrylate) (PA-18) [20–23] and maleic anhydride α-olefin alkyl ester random copolymers (MACs) [24,25]. Recently discussed examples of nucleation promoters are polyethylene–poly(ethylene propylene) diblock copolymers (PE-PEPs) [26,27] and random coplymers of poly(ethylene butene) (PEBs) [26,28]. Structures of these polymers are shown in Fig. 3.

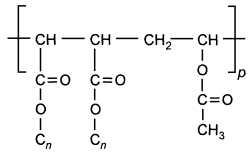
Despite their importance and a significant research effort over several decades, a complete understanding of the mechanisms by which these additives operate remains a field of active research. Many of the additives shown in Fig. 3 have hydrocarbon side chains that may interact with *n*-alkane



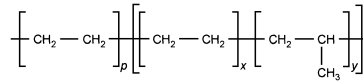
(a) ethylene-vinyl acetate random copolymer (EVA)



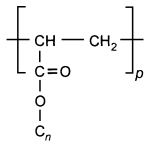
(d) maleic anhydride α-olefin alkyl ester compolymer (MAC)



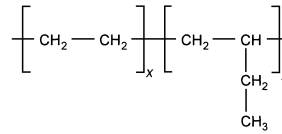
(b) fumarate-vinyl acetate alternating copolymer (FVA)



(e) polyethylene–poly(ethylene-propylene) block copolymer (PE-PEP)



(c) poly(alkylacrylate) (PA-*n*)



(f) poly(ethylene-butene) random copolymer (PEB)

Figure 3 Molecular structures of some of the nucleation promoters and growth inhibitors that have been used to modify *n*-alkane crystal growth. In these diagrams, *p* and *q* denote monomers in ordered blocks, *x* and *y* indicate randomly arranged copolymers, and *m* and *n* measure the lengths (carbon number) of alkyl chains.

molecules. There is a general consensus that the nucleation promoters are able to form aggregates in solution at temperatures above the wax appearance temperature (WAT), providing a template for *n*-alkane crystal nucleation, and that the crystallization inhibitors operate by adsorbing to crystal growth faces, blocking further growth. However, other mechanisms, such as the formation of a diffusion barrier, have been proposed. It has also been suggested that an important function of the adsorbed polymer chains is to prevent aggregation of *n*-alkane crystals, improving the flow properties of the slurry of crystals. Because of the wide variety in crude oils, even for wells in the same reservoir system [24], and the resulting variation in diesel fuels and fuel oils derived from different sources, the effectiveness of additives must be

tested for each case. A better understanding of the mechanisms by which they act enables treatments to be chosen more efficiently and confidently.

III. CRYSTALLIZATION

As a crude oil or fuel is cooled, the heavier *n*-alkanes (i.e., those with large *n*) begin to precipitate. In most cases, a range of *n*-alkanes becomes simultaneously supersaturated, resulting in solid solutions composed of several *n*-alkanes. Despite this complexity, it is useful to discuss the crystallization of pure materials. Although a complete review of crystallization theory is beyond the scope of this chapter, some discussion is necessary to understand the changes in crystallization behavior caused by additives. For comprehensive treatments, the reader is referred to such works as *The Handbook of Crystal Growth* [29] and *Crystal Growth for Beginners* [30]; here, we will restrict our attention to those aspects of crystallization that apply to the problem of additive effects.

A. Nucleation

The first step in crystallization is nucleation. A solution is said to be supersaturated if its bulk free energy is lowered by precipitation of the solute. However, the formation of crystals in the previously homogeneous solution involves the creation of an interface, which increases the free energy. The free energy is thus decreased in proportion to the number of molecules *n* in the crystal, but increased in proportion with the area of the faces:

$$\Delta G(n) = -n\Delta\mu + \sum_m \Omega_m(n)\gamma_m \tag{1}$$

where $\Delta\mu = \mu_L - \mu_S$ is the difference in chemical potential between the *bulk* liquid and solid phases, γ_m is the surface energy of the *m*th face (i.e., the crystal–solution interface), and $\Omega_m(n)$ is the area of the *m*th face. Because the areas scale as $n^{2/3}$ (assuming an energy-minimizing shape is maintained throughout nucleation), the change ΔG in free energy will always be positive for small *n*, have a maximum, ΔG^*, at some critical number of molecules n^*, and be negative for large *n*. The function $\Delta G(n)$, shown in Fig. 4, then represents a nucleation barrier: only for clusters of molecules larger than a critical size n^* is further growth favored.

In elementary treatments of classical nucleation theory, an isotropic surface energy $\gamma_m = \gamma$ is usually assumed. Because *n*-alkane crystals are decidedly anisotropic, a more realistic approximation is to assume "pancake"-shaped cylindrical nuclei with radius *r* and height *h*, as indicated in

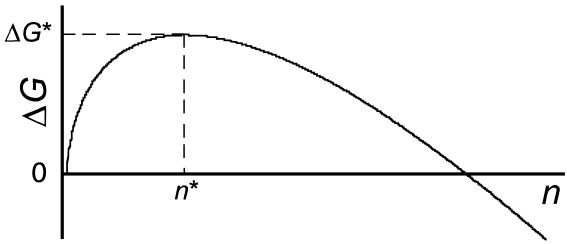


Figure 4 Free-energy change $\Delta G(n)$ relative to the fluid phase for a crystal containing n molecules. Note that only clusters containing more than the critical number n^* molecules can decrease their free energy by growing larger.

Fig. 5a, governed by an edge energy γ_e and top energy γ_0, with $\gamma_0 < \gamma_e$. The shape that minimizes the surface energy of such a crystal, given a constant volume, will have $h = 2r\gamma_0/\gamma_e$. Defining η as the ratio γ_0/γ_e (which is then also the equilibrium crystal aspect ratio), Eq. (1) takes the form

$$\Delta G(n) = -n\Delta\mu + 6\pi\left(\frac{v}{2\pi\eta}\right)^{2/3}\gamma_0 n^{2/3} \tag{2}$$

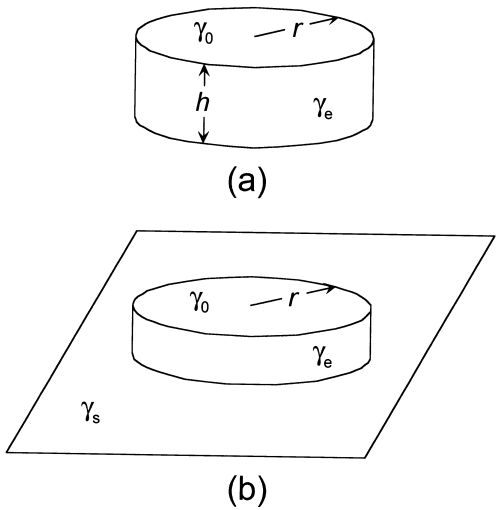


Figure 5 Nucleation of a cylindrical nucleus with top energy γ_0 and edge energy γ_e. (a) Homogeneous nucleation from solution. An aspect ratio $h/r = \gamma_0/\gamma_e$ minimizes the surface energy for a fixed volume. (b) Heterogeneous nucleation on a flat substrate with surface energy γ_s. The height, size and hence free energy ΔG^* are reduced from the homogeneous case.

where v is the molecular volume. Setting the derivative of Eq. (2) with respect to n equal to zero, we find that the maximum value of ΔG occurs at

$$n^* = 16\pi \frac{v^2 \gamma_0^3}{\eta^2 (\Delta\mu)^3} \tag{3}$$

corresponding to a critical radius of

$$r^* = \frac{2v\gamma_0}{\eta\Delta\mu} \tag{4}$$

and energy barrier of

$$\Delta G^* = 8\pi \frac{v^2 \gamma_0^3}{\eta^2 (\Delta\mu)^2} \tag{5}$$

With these definitions, we can express $\Delta G(n)$ in more compact form as

$$\Delta G(n) = \Delta G^* \left[3\left(\frac{n}{n^*}\right)^{2/3} - 2\left(\frac{n}{n^*}\right) \right] \tag{6}$$

In the approach of classical nucleation theory, the nucleation rate is expressed in terms of the number of critical nuclei per volume expected in an equilibrium system, $c \exp(-\Delta G^*/kT)$, where c is the solute concentration, the rate ω_+ at which the crystallizing species attaches to the critical nucleus, and a factor known as the Zeldovich factor, which corrects for the nonequilibrium nature of the crystallizing system [30,31]. For solution growth, the rate ω_+ is often controlled by a desolvation barrier E_{desol}, so the nucleation rate can be expressed as

$$I = K \exp\left[-\frac{(E_{\text{desol}} + \Delta G^*)}{k_B T} \right] \tag{7}$$

where k_B is Boltzmann's constant. It can be shown that the coefficient K is proportional to c^2, but only weakly dependent on the temperature.

From the form of Eq. (7), it is clear that the nucleation rate can be increased by decreasing the energy cost of the critical nucleus, either by increasing the supersaturation $\Delta\mu$ or decreasing the surface energy cost (perhaps by nucleation on an impurity surface).

B. Crystal Growth

Once a critical nucleus has formed, crystal growth can commence. To precipitate from solution, a crystallizing species must move through the solution to the crystal surface (either by diffusion or bulk fluid flow), adsorb onto the growth face (possibly involving an activated desolvation process), and

diffuse on the growth face into a preferred position and orientation on the face, thus extending the boundary of the crystal. Depending on the type of growth, any of these processes may be the rate-limiting step. For instance, if diffusion to the crystal is slow, large, well-ordered crystals may result because the timescale for ordering *on* the crystal is less than the time scale for diffusion *to* the crystal. This extreme does not guarantee, however, simply faceted crystals: as is well known, a planar interface is kinetically unstable (the Mullins–Sekerka instability [32]) and single crystals with complicated boundaries (e.g., snowflakes) are instead seen.

In many cases, the growth rate is controlled by the atomic-scale kinetics of the incorporation of adatoms with the crystal surface. An important factor is then the structure of the growth face: whether it is *faceted* or *rough*. Minimization of the surface energy favors faceted crystals terminated by low-index faces, whereas the dominance of entropy at high temperatures leads to rough interfaces. Using a simple model for a surface restricted to monatomic steps, Jackson showed that the transition between faceted and rough growth occurs at a sharply defined temperature [33–35]. He defined a parameter α, now known as the Jackson α parameter, which can be expressed in terms of the latent heat L of the phase transition, the ideal gas constant R, and the fraction ξ of nearest-neighbor bonds that lie within the lattice plane defining the crystal face:

$$\alpha = \frac{L}{RT}\xi \tag{8}$$

In this model, a value of $\alpha = 2$ marks the transition between faceted and rough growth. More realistic approaches show the same type of dependence on temperature, although the numerical value of α for the transition is altered [30].

A rough surface will contain a high density of "kink sites"—sites where an adatom can form many bonds with the growing surface. Such faces grow by direct addition of adatoms from the fluid phase—a situation called *continuous growth*—with a rate $R \sim \Delta\mu$ [30]. This is also the case for growth on facets with a high density of kink sites (so-called K faces).

Growth on a flat face with a low density of kink sites occurs by the flow of well-defined steps across the surface, maintaining a nearly faceted face. Given a step velocity v_s, step height a, and average distance between steps d, the growth rate normal to a face is $R = av_s/d$. The velocity of these steps is proportional to $\Delta\mu$ [30]. If the formation of new steps requires the nucleation of two-dimensional (2D) "islands" on the face (the "birth-and-spread" model [36]), the overall crystal growth rate is proportional to $\Delta\mu^{5/6} \exp\left(-\Delta G^*_{2D}/k_BT\right)$, where ΔG^*_{2D} is the 2D nucleation barrier [30] [in analogy with the 3D nucleation barrier described by Eq. (5)]. In many cases, the

growth rate is dominated by the presence of screw dislocations (the Burton–Cabrera–Frank or BCF model [37]). Here, no step nucleation is necessary because addition of molecules to the step merely turns the spiral. The tightness of the spiral is controlled by the critical 2D nucleation radius, which is proportional to $(\Delta\mu)^{-1}$. The overall growth rate is then proportional to $(\Delta\mu)^2$, as has been amply discussed theoretically and verified experimentally.

C. Alkane Crystals

Pure (monodisperse) *n*-alkanes with $n \geq 6$ crystallize into one of three crystal structures: orthorhombic for *n* odd and $11 \leq n \leq 43$, triclinic for *n* even and $n \leq 26$, and monoclinic for *n* even and $n \geq 26$ [38]. In all cases, the anisotropy of the long *n*-alkane molecules results in a highly anisotropic crystal structure with two of the lattice parameters nearly independent of the chain length and the third linearly dependent on *n*. For instance, in the orthorhombic structures, the unit cell has dimensions $a = 4.97$ Å, $b = 7.478$ Å, and $c = 2.546$ Å $+ (3.75$ Å$)n$ [38]. This causes a strong anisotropy in the surface energy and growth velocity, resulting in platelike crystals bounded by large {001} faces. It is the propensity of these plates to form gellike aggregates that is problematic to the petroleum industry.

The edges of the crystalline *n*-alkane plates are typically bounded by a combination of {110}, and possibly {010}, planes, depending on external parameters such as temperature and the chemical nature of the solvent. Bennema has performed an extensive analysis integrating the bond energetic approach of Hartman–Perdok theory with the theory of thermal roughening at finite temperatures to predict the equilibrium crystal form [38,39]. These studies can, for instance, explain experimental observations of different roughening transition temperatures for different faces, which can allow the slowest-growing face to remain faceted while some of the narrow faces undergo a roughening transition. This is potentially important for the design of growth modifiers that target faces with faster growth rates.

In addition to the crystalline forms, *n*-alkanes can form solids where a long-ranged order is preserved, but the orientational order about the long axis of the *n*-alkane molecules is lost. These states are known as the "rotator phases" and have received much attention recently [40]. When they exist, the rotator phases are stable in a temperature regime between those of the liquid and true crystalline states. Even at temperatures where no thermodynamically stable rotator phase exists, it has been suggested that nucleation may occur through a transient rotator phase [40,41] and that growth may proceed through a rotator precursor [39–42]. Thus, rotator phases may play a role in the interactions of additives with diesel fuels [42].

A curious phenomenon related to the formation of rotator phases is the ability of *n*-alkane liquids to form a monolayer of "crystallized" rotator phase

at the liquid–vapor interface, even at temperatures above the melting point of the bulk crystallized phase (whether rotator or solid). This phenomenon has been called "surface freezing" and occurs under conditions where the sum of the liquid–rotator and rotator–vapor surface energies is less than the liquid–vapor surface energy [43,44]. Then, the formation of a single monolayer of the rotator phase lowers the total free energy even if the temperature is above the bulk melting point of that solid. Sirota has suggested that the existence of surface-frozen layers is responsible for the anomalously small undercoolings typically observed for alkane crystallization [45]. Because these crystal surfaces may already be present when the temperature first drops below the equilibrium phase–transition temperature, nucleation is not required for crystallization to commence. Although it is unclear whether such a mechanism is important in the more complex crude oils and diesel fuels, it is possible that surface freezing will promote the growth of crystals from such systems.

When a complex fluid such as a crude oil, fuel oil, or diesel fuel is cooled, a range of *n*-alkane molecules typically becomes simultaneously supersaturated. Because these molecules are chemically similar, there is relatively little difference in their affinities to the surface of the growing crystal. For the crystal to grow with this range of molecules in the surface layers, either its lattice must distort to accommodate the different molecules, forming a solid solution, or the dissimilar molecules must be rejected into solution. It is well known that crystallization of *n*-alkanes in the presence of even small quantities of *n*-alkanes of slightly different lengths results in crystals with an orthorhombic structure, regardless of the structure that would have been obtained for the pure *n*-alkane [46].

IV. ADDITIVES

There are several mechanisms by which an additive can affect a crystallizing system. For instance, the often-cited melting-point depression is a thermodynamic effect arising from increase in entropy of the fluid phase due to the presence of a solvated impurity. We will be discussing additives that have a minimal effect on the thermodynamics, but a dramatic effect on the kinetics, of crystallization. Some additives act to increase nucleation rates, in which case they are known as *nucleation promoters*, whereas others retard the growth rate and are termed *kinetic inhibitors*.

A. Nucleation Promoters

The nucleation barrier is always decreased if nucleation can occur on a foreign surface. As shown in Fig. 5b, nucleation of a cylindrical crystal on a surface replaces one crystal surface of area A and surface energy γ_0 by an interface of energy $A\gamma_i$, where γ_i is the interfacial energy per area of the new interface

created between the crystal and foreign surface. In the process, the surface energy of the foreign surface is decreased by $A\gamma_s$, where γ_s is its surface energy. For this case of heterogeneous nucleation, all faces of the critical nucleus which do not contact the foreign surface have the same shape and size, but the height normal to the surface and the nucleation barrier are altered [30]. Commercial nucleation promoters for the *n*-alkanes are thought to operate by providing such a template upon which the crystal can nucleate. For our pancake model, the nucleation barrier for heterogeneous nucleation becomes

$$\Delta G^*_{\text{het}} = \Delta G^* \left(\frac{\gamma_0 + \gamma_i - \gamma_s}{2\gamma_0} \right) = \Delta G^* \left(1 - \frac{\Gamma}{2\gamma_0} \right) \tag{9}$$

where $\Gamma = \gamma_0 + \gamma_s - \gamma_i$ is the adhesion energy between the crystal and substrate.

For *n*-alkane nucleation promoters, this template is thought be formed by the aggregation of the polymeric additives into surfaces similar in structure to *n*-alkane crystals [42], as is shown schematically in Fig. 6. In the limiting case that the surface of these aggregates forms a structure identical to that of the *n*-alkane crystal, so that the interfacial surface energy γ_i vanishes and $\Gamma \to 2\gamma_0$, the nucleation barrier vanishes. Then, the template serves as a seed crystal and further crystallization becomes a problem in crystal growth only.

B. Kinetic Inhibitors

Kinetic inhibitors alter the kinetics of growth by adsorbing to growth faces and "poisoning" sites necessary for growth. Cabrera and Vermilyea first

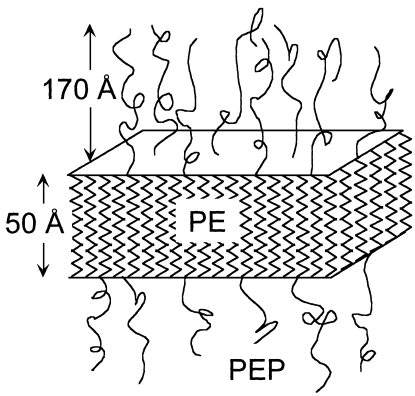


Figure 6 Self-assembled polymer aggregate consisting of a crystalline polyethylene (PE) core surrounded by disordered poly(ethylene propylene) brushes. The PE surfaces provide a template for *n*-alkane nucleation. (Adapted from Ref. 27.)

described the basic mechanism whereby this might occur [47]. In the case of faceted growth, where the crystal grows via step flow, a kinetic inhibitor may alter the crystallization rate by adsorbing to the growth face and slowing the passage of steps. If the impurity sites pin the step, as shown in Fig. 7, the step must curve as it passes through the "fence" of impurities. The step can detach from the impurity fence when it attains a radius of curvature at least as small as half of the distance between the pinning sites. At that point, the curved segments on either side of a pinning site can merge, allowing the step to

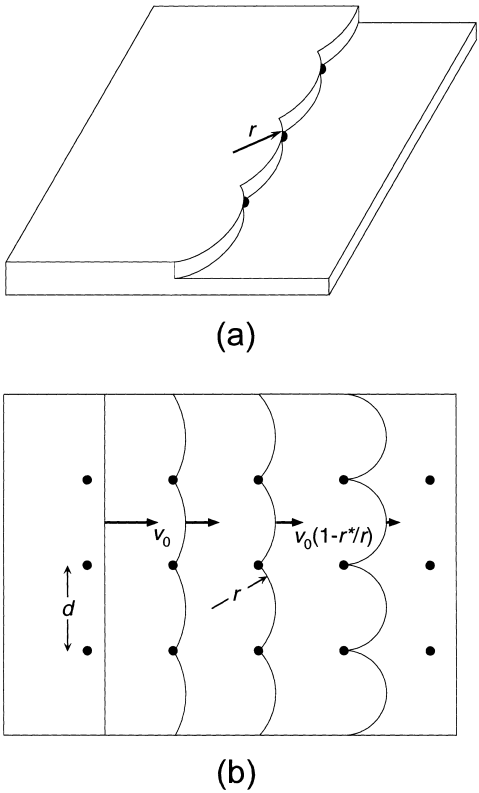


Figure 7 Step pinning by absorbed impurities. (a) Oblique view of a single step pinned by impurities (black disks). The radius of curvature of a portion of the step between pinning sites is r. (b) Top view of multiple steps (solid curves) in various stages of proceeding through a regular array of impurity sites (black disks). The step velocities depend on the instantaneous radius of curvature and vary between the velocity of the straight step, v_0, to $v_0(1 - 2r^*/d)$ when the step just breaks free of a row of pinning sites.

advance. According to Laplace's law (in two dimensions), the curvature causes a reduction in the supersaturation:

$$\Delta\mu(r) = \Delta\mu_\infty - \frac{\beta\Omega}{r} \tag{10}$$

where $\Delta\mu_\infty = \mu_L - \mu_S$ is the supersaturation of a straight step, β is the energy cost per length of step, Ω is the molecular area of the crystallizing species (projected onto the growth face), and r is the radius of curvature of the step. Because the step velocity is proportional to the local supersaturation $\Delta\mu(r)$, this leads to a decreased velocity of

$$v(r) = v_0\left(1 - \frac{r^*}{r}\right) \tag{11}$$

where v_0 is the velocity of a straight step and $r^* = \beta\Omega/\Delta\mu$ is the critical nucleus of two-dimensional islands on the surface [30].

This model predicts that if the pinning sites have a typical spacing d less than twice the critical radius, forcing steps proceeding through the impurity fence to have a radius of curvature smaller than r^*, growth will cease altogether. A critical supersaturation

$$\Delta\mu_X > \frac{2\beta\Omega}{d} \tag{12}$$

required for growth to occur can then be defined. Because the critical radius is a decreasing function of supersaturation, growth of a poisoned surface can resume if the supersaturation is increased. These predictions have been borne out by experiments which show a negligible growth rate in the presence of inhibitors, even if the system has significant undercooling or supersaturation [48]. An important quantity is the average spacing of additives, which will depend on the adsorption isotherm and the adsorption kinetics (which may be quite slow for polymeric additives). If the minimum separation distance of additives on the surface (determined by either the crystal lattice or the size of the impurity molecule) is b and the additives cover a fraction X of the available sites, the average separation between pinning sites will be (within numerical factors) $b/\sqrt{X}$. The condition for completely arresting the growth then becomes

$$\Delta\mu_\infty < \frac{2\beta\Omega}{b}\sqrt{X} \tag{13}$$

Assuming that the coverage is in equilibrium and follows the Langmuir isotherm

$$X = \frac{Kc_X}{1 + Kc_X} \tag{14}$$

where c_X is the additive concentration in solution and K is the equilibrium constant for the adsorption process [49], the maximum supersaturation for which growth can be arrested depends on the additive concentration as $\sqrt{Kc_X/1 + Kc_X}$. Similar dependences on additive concentrations are seen experimentally for the inhibition of *n*-alkane crystallization [50], providing indirect evidence that these additives act via surface adsorption. Because the additives used in these cases are polymeric, adsorbed polymer chains can be expected to have a minimum separation b of the order of the polymer radius of gyration R_g.

Although it is clear that the crystal growth rate can be decreased by this mechanism, the precise form of this decrease has been disputed. Assuming a regular lattice of pinning sites as shown in Fig. 7b, it is clear that the step velocity will vary between v_0 and $v_0(1 - 2r^*/d)$. One might arbitrarily take either the arithmetic or the geometric mean of the two extreme velocities to obtain effective velocities of either

$$v_{\text{arith}} = v_0\left(1 - \frac{\beta\Omega}{\Delta\mu}\frac{\sqrt{X}}{b}\right) \tag{15}$$

or

$$v_{\text{geom}} = v_0\left(1 - \frac{2\beta\Omega}{\Delta\mu}\frac{\sqrt{X}}{b}\right)^{1/2} \tag{16}$$

respectively [51], although there is no strong theoretical justification for these forms. Alternate discussions of the effect of additives on growth velocity instead calculate the fraction of steps that contain unpoisoned segments longer than $2r^*$ and, hence, are able to advance [52,53].

Complicating the situation somewhat is the expectation that the impurities would not normally be arranged in the ordered array shown in Fig. 7b. Potepenko has treated the problem of step flow through a random fence of impurities as a percolation problem, whereby growth can proceed if there exists a path across the growth face following which a step need never have a radius of curvature that is smaller than the critical nucleation radius [54,55]. In this case, complete inhibition can still occur, but at a different adsorption level than would be predicted for an ordered array of impurities.

Because the mechanisms for both nucleation promoters and growth inhibitors involve interactions between the additive and crystal surface, the same molecule may serve a dual purpose. Any molecule that can nucleate crystals may also be able to adsorb to the growth faces of an already existing crystal, modifying its growth rate. Studies have shown that in some cases the dependence of growth rate and crystal size is a nonmonotonic

function of additive concentration, suggesting the existence of two additive mechanisms [56].

Such mechanisms for the additive activity have not been the only ones proposed. Davey has argued that the effective growth rates suggested by Eqs. (15) and (16) contradict experimental evidence and suggests that instead, the equation

$$v_{\text{diff}} = v_0(1 - X) = \frac{v_0}{1 + Kc_X} \tag{17}$$

should be used [57]. In this model, it is assumed that the main function of the adsorbed inhibitors is to decrease the rate of surface diffusion of crystallizing molecules to steps. However, this may not apply to solution growth, where bulk diffusion of the crystallizing species through solution to the steps usually dominates over surface diffusion [30]. On the other hand, a layer of adsorbed polymeric inhibitor may decrease the rate of heat and mass diffusion across the interface, both increasing the supersaturation necessary for growth and raising the effective dissolution point [21]. For the case of *n*-alkane crystallization, it has also been suggested that an important contribution to the improved rheological properties induced by polymeric additives is the ability of the adsorbed polymer layer to prevent aggregation [58], perhaps via an entropic pressure induced by confinement of polymer chains.

If growth occurs on a face with a high density of kink sites (either due to its crystallographic orientation or to a growth temperature above the roughening transition), the above mechanism for inhibition clearly cannot apply. Van Enckevort and van den Berg have noted that a spherical bulge growing between pinning sites can always advance, but they suggest that growth can be arrested by the increase in surface energy due to the adsorbed additives, which lowers the effective supersaturation [59].

Although these models generally assume that the kinetic inhibitors are irreversibly adsorbed, this is not a necessary requirement for an inhibitory effect. If the inhibitors can desorb from or diffuse on the surface, slow growth will still be possible, even at high impurity coverage. In this case, growth is controlled by the activation energy of breaking bonds between the additive and surface [59]. Even slow desorption will allow a "completely" poisoned system to eventually attain equilibrium. For the remediation of *n*-alkane crystallization from diesel fuels, it is often sufficient that this process take significantly longer than 1 day.

C. Control of Growth Shape

In addition to decreasing the rate of crystal growth at a given driving force, kinetic inhibitors have the potential to control the crystal habit by differential

adsorption to different growth faces [60]. For instance, the lead ion Pb^{2+} alters the crystal morphology of NaCl from its usual cubic form to an octahedral form by slowing the growth of the {111} faces until they become the limiting face [61].

Growth modifiers have long been used in industry to control crystal habit. A typical growth inhibitor designed to produce a particular growth habit will consist of two subunits: one that readily adsorbs to a desired growth face and one that prevents further growth on that face, making it the slowest growth face [39,60,62,63]. This is shown schematically in Fig. 8.

This mechanism may be important for the inhibition of *n*-alkane crystallization. The additives in typical use (see Fig. 3) contain hydrocarbon groups that presumably adsorb to growth faces, connected to a polymeric chain that cannot be incorporated into the crystal structure. Effective additives should target the plate edges (the fastest-growing surfaces), which are often thermally roughened. In this case, the strong anisotropy in the surface energy should lead to the growth of smaller plates between additive sites on the plate edges. If the reduction of the growth rate on the edges is greater than that on the larger plate surfaces, the aspect ratio of the resulting crystals will be decreased, possibly improving the low-temperature flow properties of the remaining fluid. Optical studies of the shape of *n*-alkane crystals in the presence of growth modifiers

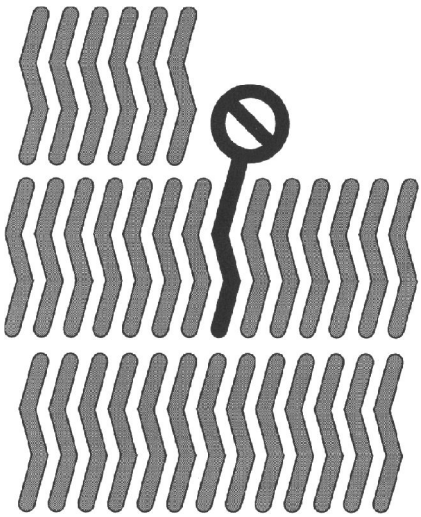


Figure 8 Schematic of a "tailor-made" additive (shaded black) adsorbed to a specific growth face, blocking step growth.

show that they become more rodlike (in addition to smaller), with changes observed at additive concentrations as low as 1 ppm [42].

D. Examples of Crystallization Inhibition

Examples of crystallization inhibition are abundant in the literature. Although the purpose of this chapter is not to review the history of kinetic inhibitor studies in general (the reader is instead referred to other, more general reviews [53,57,62–64]), a few representative examples are listed for comparison with additive activities in diesel fuels.

A common use of growth modifiers is to prevent crystallization of undesired solids. For instance, scale inhibitors are used in both water transport systems [65,66] and refrigeration coolant loops [67] to prevent the formation of potential blockages. In a more extreme example, growth modifiers have been used to completely arrest an undesired crystal in a complex system, allowing the formation of a desired product with a free energy larger than that of the arrested crystal [68].

One example that is quite similar to the problem discussed in this chapter is the prevention of hydrate crystallization in natural gas pipelines. Hydrates are molecular crystals consisting of water and a small "guest" molecule (such as methane) [69]. Natural gas hydrates can be found in a pressure–temperature range that typically occurs on continental shelves. The resource represented by such deposits (it is estimated that the amount of natural gas in hydrates exceeds the combined amount of all known conventional methane reserves) and its global climate implications (an increase in ocean temperatures could cause methane—a greenhouse gas—to be released from hydrates) have been discussed recently [70]. The petroleum industry has devoted significant effort into preventing the formation of hydrates in natural gas pipelines. Conventional thermodynamic approaches, such as insulating the pipeline to maintain high temperature and adding methanol to shift the thermodynamic formation temperature, are both costly and potentially environmentally dangerous, so growth modifiers that can inhibit crystallization for the residence time of the components in the pipeline have begun to be used [71–74].

Kinetic inhibitors are also used in industry to control crystal shapes. For instance, they are used in the plaster industry to control the shape of gypsum crystals and, hence, the mechanical strength of the resulting solids [75]. Such control is also important in the food and pharmaceutical industries, where the shape of grains in powders influences their flow and mixing properties [42].

Nature, too, employs additives to control biomineralization. In some cases, it is important that a particular material *never* crystallize. For instance, certain species of Arctic fish live their entire lives at temperatures below the freezing point of their bodily fluids without freezing [76,77]. Proteins (termed

"antifreeze glycoproteins") found in these fish apparently adsorb quickly to the faces of any ice nuclei, preventing their further growth. Similarly, naturally occurring proteins arrest the growth of calcium oxalate in kidneys, inhibiting the formation of kidney stones [78,79].

The precise control of crystallization possible by additives designed by nature is demonstrated by biomineralization of structures such as seashells, bone, and teeth [80–82]. Proteins extracted from the surfaces of natural calcium carbonate structures such as oyster shells have been demonstrated to modify growth of calcium carbonate solutions, even in the absence of a living organism [83].

V. STUDIES OF *n*-ALKANE INHIBITORS

Additives in the petroleum industry are often classified as nucleation promoters or as crystallization inhibitors depending on their operational effects. It is often assumed that nucleation promoters will result in many small crystals, improving flow through filters and improving performance in flow tests, whereas crystallization inhibitors will decrease crystal growth rates, delaying the appearance of crystals of significant size. However, the observation that different studies will sometimes ascribe different behaviors to the same additive emphasizes the need to classify additives according to their microscopic mechanisms. Because the additives interact in all cases with *n*-alkane crystals (i.e., either by aggregating into templates upon which protonuclei can form or by adsorbing to and poisoning growth faces), it is possible for the same additive to serve a dual role. In fact, it is likely that any polymer with alkyl chains able to self-assemble into a nucleation template can also adsorb individually onto growth faces under the right conditions. Evidence for such dual roles has been reported [42,56].

Studies of the effect of additives on wax growth from crude oils and diesel fuels are often difficult to interpret due to the complexity of the system. This difficulty leaves researchers with two choices: (1) to study the real, complex system in order to obtain reliable data of commercial importance or (2) to study simpler model systems in the hope that insights gained might be transferable to the systems of interest. Both approaches have merit as well as potential pitfalls. Data obtained for a real system may not be easily transferable to other similar systems, forcing experiments to be repeated for every fuel formulation. Results from well-characterized model systems may be irrelevant to commercially important systems due to the omission of important components. Intermediate approaches have also been used (e.g., the doping of crude oils by specific *n*-alkane molecules) [25].

Our present understanding of the effect of additives on *n*-alkane crystallization is the result of three decades of research, not all of which is available in

the open literature. In the following subsections, a few of the studies that have improved our knowledge of these systems are highlighted.

A. Standard Tests

The flow properties of a crude oil determine its ability to be delivered through a pipeline. Similarly, operation of a diesel engine relies on delivery of the fuel through fuel lines, filters, and injectors. The appearance of *n*-alkane crystals as the temperature is lowered can prevent such operations. This creates the need for tests that predict the low-temperature behavior. The most reliable test of low-temperature operation is the operation itself. In most cases, it is desirable to perform a less costly test in advance of system failure. One approach is to mimic the relevant system as closely as possible. For instance, the operability of diesel engines can be tested by cooling an entire vehicle while monitoring fuel delivery and engine operation [11]. Because such large-scale experiments are quite inconvenient, one often tests properties in smaller systems.

Standard tests monitor the appearance of wax crystals and their effects on flow properties in small sample volumes as the temperature is lowered. For instance, the cloud point (CP) temperature is measured by cooling 45 mL of fluid until cloudiness in the fluid can be discerned [1]. The wax appearance temperature (WAT) similarly measures the temperature of the first appearance of significant quantities of *n*-alkane crystals when a sample is cooled while being stirred [84]. The viscosity of a crude oil or fuel increases dramatically with increasing crystallization of the *n*-alkanes. The pour point (PP) measures the temperature below which a 45-mL sample is too viscous to visibly flow under its own weight [3].

Most diesel engines and systems that rely on similar fuel oils are able to operate at temperatures below the WAT and CP, but fail at temperatures above the PP. Flow tests allow better predictions of low-temperature performance; thus, fuel standards are generally specified in terms of minimum flow-test performance (which may vary with location and season). The cold-filter plug point (CFPP) is the temperature at which 20 mL of a sample will no longer flow through a 45-μm filter with the aid of a 2-kPa pressure difference [2]. The more rigorous low-temperature flow test (LTFT) measures the minimum temperature at which 180 mL of fluid will flow through a 17-μm filter in 60 s or less with the aid of a pressure of 20 kPa [4]. The LTFT is used to set minimum flow performance standards in North America.

The ability of additives to improve the low-temperature performance of a crude oil or fuel is typically reported in terms of a reduction in the temperatures returned by these standard tests. For instance, the CP can typically be lowered by $\sim 3^{\circ}C$ by the addition of 500 ppm of additive, although the improvement saturates for much higher additive concentration [50].

Although flow tests are a practical method of assessing fuel performance and performance gains due to additives, they do not provide additional information about the additive mechanisms. For the development of improved additives, information provided by the standard tests must be correlated with data obtained from other studies.

B. Rheological Testing of the Fluids

Because it is ultimately the flow of crude and fuel oils that is commercially important, significant effort has been devoted to understanding the modification of rheological properties by additives. These effects have been investigated as a function of additive concentration [58], demonstrating not only the significant gains that can be realized by the addition of small quantities of additives but also that the improvement saturates at finite additive concentration [50]. Such a saturation is expected of any effect relying on surface adsorption. Rheological testing as a function of concentration and the vinyl acetate content of random EVA copolymers has shown that additives only modify the viscosity of crude oils at temperatures below the WAT [15]. This provides clear evidence that such additives modify the nature of the *n*-alkane crystals, rather than the rheological properties of the uncrystallized fluid.

C. Studies of the *n*-Alkane Crystals

The form of *n*-alkane crystals and crystal networks can be studied by optical microscopy. The decrease in crystal size and change in crystal habit have been demonstrated for both commercial [5] and model systems [38]. Typically, growth inhibitors are able to decrease *n*-alkane crystal sizes in fuels from 60 to 10 μm, whereas nucleation promotors are able to achieve sizes as small as 1 μm [5]. The greater reduction by the nucleation promoters is presumably due to the creation of many nuclei so that the crystallized fraction is distributed among many small crystals.

The crystal habits have also been studied at higher resolution by scanning electron microscopy (SEM). Van Rosmalen and Bennema have reported that, in some cases, additives cause the *n*-alkane crystals to form needles bounded by {111} faces, indicating that the additives adsorb to those faces [60]. Surprisingly, only four of the eight {111} faces became large. This symmetry breaking is difficult to explain on the basis of the crystal structure, although it is interesting to note that similar behavior has been seen for hydrate crystals in the presence growth inhibitors [71,73]. In the case of hydrate crystals, it was suggested that a kinetic symmetry breaking occurs during growth.

X-ray crystallography has shown that *n*-alkane crystals grown from diesel fuels retain an orthorhombic structure in the presence of growth inhibitors, consistent with a model involving surface adsorption of additives [5,42]. The

former study also showed that the much smaller crystals formed in the presence of nucleation promoters instead had a hexagonal structure, which the authors ascribed to a finite-crystal-size effect, rather than a change in the bulk equilibrium structure. It seems likely that they were measuring a rotator-phase structure.

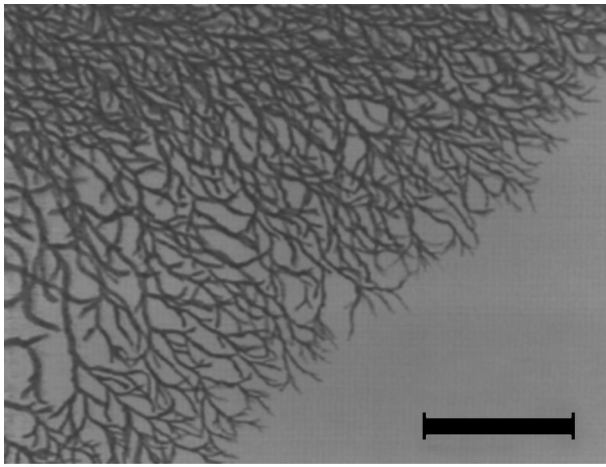
Little attention has been devoted to the structure of wax crystals on a scale much larger than the individual crystals. My research group is using optical microscopy to study the crystal morphology obtained from model systems composed of tricosane ($C_{23}H_{48}$) dissolved in dodecane ($C_{12}H_{26}$) on a scale of hundreds of microns. Figure 9 shows the crystallization of such a system (of 75 μm thickness) in the presence of additives such as (Fig. 9a) EVA and (Fig. 9b) PA-18. We find branched structures that are apparently connected over large distances (Hutter, unpublished data). It is interesting to note that EVA, which is usually considered to be a nucleation promoter, has a large effect on the growth form (compare with Fig. 2).

D. Nucleation Studies

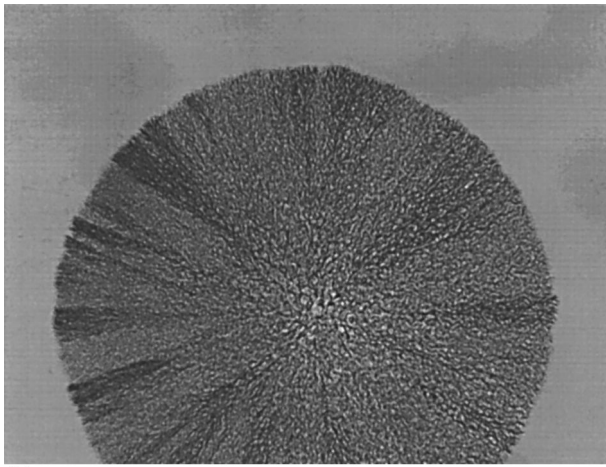
Nucleation studies are, in general, difficult due to the small size of the nucleus and the interference of growing crystals after the first nuclei form. Such studies are, however, indispensable for a proper understanding of the effect of nucleation promoters.

One way to separate nucleation rates from growth rates is to study emulsified systems. In this way, nucleation leads quickly to solidification of an entire droplet, but no further bulk growth occurs [85]. The nucleation rate can then be easily measured by any technique that can measure the total fraction of solidified sample. For instance, differential scanning calorimetry (DSC) has been used to measure homogeneous nucleation rates of pure *n*-alkanes and, hence, their surface energies (given assumptions about the nucleus shape) [86]. This approach could be extended to obtain similar data in the presence of nucleation promoters.

It is believed that nucleation promoters self-assemble into ordered structures prior to crystallization, providing a template for *n*-alkane nucleation [42]. The structure of these templates in solution has been studied by small-angle neutron scattering (SANS). Studies of PE–PEPs, additives known to effect nucleation rates, indicate that they form platelike structures with ordered cores consisting of polyethylene blocks 20–100 Å thick surrounded by amorphous poly(ethylene propylene) brushes extending to approximately 200 Å from the cores [26,27] (see Fig. 6). Further studies using the contrast-variation technique, whereby neutron scattering contrast between specific components is maximized, revealed that the brushes elongate in the presence of *n*-alkanes at low temperatures, suggesting cocrystallization with the *n*-alkane



(a)



(b)

Figure 9 Optical micrograph of a thin sample of $C_{23}H_{48}$ crystals growing isothermally from a 60% solution with $C_{12}H_{26}$. (a) Growth with 1% EVA. Note the highly branched connected network of needle crystals. (b) Growth with 1% PA-18. Note the branched network of fine crystals. The scale bar, shown in (a), measures 100 μm.

molecules. Similar results were seen for the structure of PEB aggregates in the presence of *n*-alkanes [28]. These findings were interpreted as evidence that polymer aggregates act as a template for nucleation of wax plates, likely significantly decreasing the nucleation barrier.

E. Crystallization Rate Studies

Crystallization rates can be measured by any technique that can monitor the rate of change of crystallized volume. For instance, the overall rate of crystallization in a sample volume can be monitored thermally (as the heat evolved during the crystallization process) and by X-ray diffraction (as the increase in intensity of a crystallographic peak). Experiments have shown good consistency between methods [87].

Differential scanning microscopy has been used to monitor the nucleation [20] and growth [22] of *n*-alkane crystals in the presence of additives. It has been suggested that DSC analyses can provide more reliable estimates of fuel operability than can standard tests such as the CFPP test and LTFT [88]. Such analyses have also been employed to study the response of alternative diesel fuels derived from vegetable oils to additives [89,90].

F. Effects of the Solvent

Although additives specifically designed to adsorb to crystal faces will clearly alter crystallization kinetics, less specific binding can also have important effects. For instance, the solvent is important, both in its interactions with solvated species (changing the desolvation barrier) and with the growth face itself (changing the interfacial energies) [91]. Such effects are seen in *n*-alkanes. For example, Lin and Bennema have reported that tetracosane (n-$C_{24}H_{50}$) grown from iso-octane solutions exhibits needlelike shapes, rather than the flat plates normally seen in the absence of additives [39].

Because the solvent can influence the surfaces of *n*-alkane crystals, it may also influence the effectiveness of additives. For instance, the FVA copolymers show little activity for model fuels dissolved in *n*-heptane, in contrast to their strong activity in *m*-xylene-based model fuels [19]. Thus, fuels derived from different sources can be expected to differ in their response to treatment, even if they have similar distributions of heavy *n*-alkanes.

G. Effects of the *n*-Alkane Mass Distribution

When a range of *n*-alkane molecules is present, crystallization can be modified in nontrivial ways. For instance, the nucleation rates of eicosane (n-$C_{20}H_{42}$) and tetracosane (n-$C_{24}H_{50}$) are generally decreased by the presence of other

n-alkane molecules with similar chain lengths [92,93]. In these cases, it was suggested that the impurity molecules are incorporated in the crystal, creating a lattice mismatch that raises the nucleation barrier. Similar studies have shown that growth rates can be reduced and even halted on some faces [94]. These studies showed the {001} faces to be slowed by more than the {110} faces, resulting in increased aspect ratios of the already platelike crystals.

As a more extreme example, experiments with monodisperse, very long alkane molecules (e.g., *n*'s of a few hundred) have shown a peculiar form a self-poisoning, leading to a minimum in the growth rate as the temperature is lowered. Such molecules can crystallize into either an extended-chain or folded-chain conformation, with the folded chain being favored at lower temperatures. In the crossover regime, folded molecules can adsorb to a crystal that is growing from extended molecules, preventing further growth until the adsorbed molecule either desorbs or unfolds [95,96].

Given the above observations, it is not surprising that the distribution of *n*-alkane molecules in crude and fuel oils affects both their crystallization rates and the effectiveness of additives. Fuels with a narrow distribution are considered to be difficult to treat because *n*-alkane precipitation will then have a sharp onset as the temperature is lowered [8]. When a broader range is present, the heavier *n*-alkanes can precipitate in advance of the lighter molecules, forming templates upon which the lighter molecules precipitate as the temperature is lowered further [9]. Reddy has defined a simple figure of merit for diesel fuels: The *response number* is the negative reciprocal of the slope of the mass distribution (in mass %) of *n*-alkanes as a function of *n*, for *n* larger than the peak [7,11]. He noted that fuels with larger response numbers, corresponding to a longer high-*n* tail in the distribution, are easier to treat. Simple dilution with lighter fuels will increase the response number (by reducing the height of the *n*-alkane peak and therefore the magnitude of the slope at large *n*), suggesting that the diluted fuels will be easier to treat by additives. Such a synergistic effect has been observed for difficult-to-treat fuels, where combined dilution and additive treatment resulted in decreases to the CFPP larger than the sum of decreases obtained by either method individually [11]. However, this picture cannot be complete because studies of crude oils have shown that the additive effectiveness is *decreased* by deliberate addition of large alkanes, even though the response number would presumably be increased [25].

Attempts have been made to model the phase behavior of diesel fuels. Recent results have show good agreement between the predicted phase behavior and measured *n*-alkane distributions in the liquid and solid phases of model [97,98] and diesel [10] fuels. Clearly, such studies are an important step toward predicting the interactions of additives with complex and varying fuel compositions.

H. Additive Properties

The observation that nucleation and growth in model *n*-alkane systems can be affected by the addition of *n*-alkanes of different chain lengths suggests that the ability of additives to adsorb to growing crystals may depend on the length of their side chains. By studying a series of a FVA copolymers in model fuels consisting of *n*-dotriacontane (n-$C_{32}H_{66}$) in *m*-xylene, Beiny et al., have shown that the FVA effectiveness depends strongly on the length of their alkyl side chains, showing a maximum effectiveness at lengths near that of the crystallizing *n*-alkane [19]. FVA copolymers with side chains smaller than $n \sim 8$ exhibited no effect on crystal growth rates [19] and, in other studies, no evidence of surface adsorption [17].

There is some evidence that additives can effect the structure of solutions at temperature above the WAT. This is certainly the case in the self-assembled nucleation templates and may be more widely occurring. Some evidence of association between EVA random copolymers and *n*-alkane molecules at temperatures above the cloud point has been provided by Davidson and Cameron, who performed electron spin resonance (ESR) studies of labeled additives [16]. Machado et al., have further noted that EVA compositions that precipitate at temperatures significantly higher than the WAT have reduced effect on *n*-alkane crystallization [15], suggesting that the solubility of additives plays an important role. However, similar ESR studies of the activity of FVA copolymers (a growth inhibitor) show no change in the polymer conformation at temperatures above the CP, when *n*-alkane crystals first appear [17,18]. These additives must, therefore, only act on crystals that have already nucleated.

VI. CONCLUSIONS

Additives that are able to either increase nucleation rates or inhibit growth rates have been used by the petroleum industry to prevent the formation of crystalline networks in crude oils, diesel fuels, and other fuel oils. Although the effectiveness of such additives has been amply demonstrated both in the field and in the laboratory, the microscopic details of the mechanisms by which they act are not yet completely understood. Such information is vital to build the predictive models made necessary by the varying fuel compositions, which are driven by economic need and changing crude oil sources.

Better molecular-scale data will be critical for the creation and verification of models for additive mechanisms. Already, small-angle neutron scattering has shown promise as a means to gather such information. However, the interpretation of SANS results is model dependent, resulting in some ambiguity in the results. Real-space techniques, such as atomic force microscopy,

may make it possible to resolve some of these difficulties and directly show the location and distribution of the inhibitor molecules. Already, this tool has been used to study step structures in the presence of additives in calcite [48,99] and potassium dihydrogen phosphate (KDP) [100]. The latter study suggested that the recovery of growth (due to an increase in supersaturation) of a poisoned surface is due to the motion of macrosteps, not simply the renewed ability of steps to proceed through the impurity fence, as would be predicted by simple models. Such microscopic information is crucial to an improved understanding of inhibition mechanisms.

Numerical methods have been employed to predict macroscopic aspects of *n*-alkane crystallization, such as the equilibrium crystal structure and phase diagram. With the increasing power of numerical methods, insight may be gained from molecular-scale simulations of the interaction of additives with a growing crystal, allowing the theory of additive effects to progress beyond mean-field models. Duffy and Rodger have recently reported molecular dynamics simulations of the inhibition of octacosane ($C_{28}H_{58}$) growth by PA-18. They suggest that *n*-alkanes adsorbing on an inhibited alkane surface form a layer on top of the inhibitor molecules that is incommensurate with the underlying surface, resulting in defects that block further growth [101].

Although the cold-weather precipitation of *n*-alkanes from fuels represents a very specific technological problem, it has much in common with other systems of industrial and scientific interest. The benefits of an improved understanding of kinetic inhibitors will therefore extend beyond the needs of the petroleum industry to techniques used to control crystallization in other commercial processes, as well as to mechanisms of biomineralization at work in living organisms.

REFERENCES

1. Annual Book of ASTM Standards 5.01:D 2500-99, 2002.
2. Annual Book of ASTM Standards 5.04:D 6371-99, 2002.
3. Annual Book of ASTM Standards 5.01:D 97-96a, 2002.
4. Annual Book of ASTM Standards 5.02:D 4539-98, 2002.
5. Botros, M.G. Society of Automotive Engineers, Inc. 1997, Report No. 972899.
6. Rossemyr, L.I. Ind. Eng. Chem. Prod. Res. Dev. 1979, *18*, 227–230.
7. Reddy, S.R.; McMillan, M.L. Society of Automotive Engineers, Inc., 1982, Report No. 811181.
8. Knepper, J.I.; Hutton, R.P. Hydrocarbon Process 1975, *54*, 129–136.
9. Zielinski, J.; Rossi, F.; Stevens, A. Society of Automotive Engineers, Inc. 1985, Report No. 841352.
10. Coutinho, J.A.P.; Dauphin, C.; Daridon, J.L. Fuel 2000, *79*, 607–616.
11. Reddy, S.R. Society of Automotive Engineers, Inc. 1985, Report No. 841351.

12. Manka, J.S.; Sopko, T.M. Society of Automotive Engineers, Inc. 1998, Report No. 982575.
13. Humphrey, T. Int. J. Vehicle Design 1985, *6*, 469–474.
14. Holder, G.A.; Winkler, J. Nature 1965, *207*, 719–721.
15. Machado, A.L.C.; Lucas, E.F.; González, G. J. Petrol. Sci. Eng. 2001, *32*, 159–165.
16. Davidson, I.G.; Cameron, G.G. Macromolecules 1995, *28*, 5964–5966.
17. Davidson, I.G.; Cameron, G.G. Polym. Int. 1994, *34*, 443–447.
18. Davidson, I.G.; Cameron, G.G. Polym. Int. 1994, *34*, 449–451.
19. Beiny, D.H.M.; Mullin, J.W.; Lewtas, K. J. Cryst. Growth 1990, *102*, 801–806.
20. Ding, X.; Qi, G.; Yang, S. Polymer 1999, *40*, 4139–4142.
21. Colussi, A.J.; Hoffmann, M.R.; Tang, Y. Langmuir 2000, *16*, 2405–2408.
22. Kök, M.; Letoffe, J.M.; Claudy, P.; Martin, D.; Garcin, M.; Volle, J.L. J. Thermal Anal. 1997, *49*, 727–736.
23. Giorgio, S.; Kern, R. J. Cryst. Growth 1983, *62*, 360–374.
24. García, M.d.C.; Carbognani, L.; Orea, M.; Urbina, A. J. Petrol. Sci. Eng. 2000, *25*, 99–105.
25. García, M.d.C. Energy Fuels 2000, *14*, 1043–1048.
26. Schwahn, D.; Richter, D.; Wright, P.J.; Symon, C.; Fetters, L.J.; Lin, M. Macromolecules 2002, *35*, 861–870.
27. Leube, W.; Monkenbusch, M.; Schneiders, D.; Richter, D.; Adamson, D.; Fetters, L.; Douis, P.; Lovegrove, R. Energy Fuels 2000, *14*, 419–430.
28. Schwahn, D.; Richter, D.; Lin, M.; Fetters, L.J. Macromolecules 2002, *35*, 3762–3768.
29. Hurle, D.T.J., Ed.; *Handbook of Crystal Growth*; Elsevier Science Publishers: Amsterdam, 1993.
30. Markov, I.V. *Crystal Growth for Beginners*; World Scientific: Singapore, 1995.
31. Becker, R.; Döring, W. Ann. Phys. 1935, *24*, 719–752.
32. Mullins, W.W.; Sekerka, R.F. J. Appl. Phys. 1964, *35*, 444–451.
33. Jackson, K.A. Mater. Sci. Eng. 1984, *65*, 7–13.
34. Jackson, K.A. *Growth and Perfection of Crystals*; Doremus, R.H., Roberts, B.W., Turnbull, J., Eds.; Wiley: New York, 1958.
35. Jackson, K.A.; Uhlmann, D.R.; Hunt, J.D. J. Cryst. Growth 1967, *1*, 1–36.
36. van Leeuwen, C.; van der Eerden, J.P. Surface Sci. 1977, *64*, 237–250.
37. Burton, W.K.; Cabrera, N.; Frank, F.C. Phil. Trans. Roy. Soc. (London) 1951, *243A*, 299–358.
38. Bennema, P. *Handbook of Crystal Growth*; Hurle, D.T.J., Ed.; Elsevier Science Publishers: Amsterdam, 1993; 477–579.
39. Liu, X.-Y.; Bennema, P. Phys. Rev. B 1994, *49*, 765–775.
40. Sirota, E.B.; King, H.E. Jr.; Singer, D.M.; Shao, H.H. J. Chem. Phys. 1993, *98*, 5809–5824.
41. Sirota, E.B.; Herhold, A.B. Science 1999, *283*, 529–532.
42. Kern, R.; Dassonville, R. J. Cryst. Growth 1992, *116*, 191–203.
43. Wu, X.Z.; Ocko, B.M.; Sirota, E.B.; Sinha, S.K.; Deutsch, M.; Cao, B.H.; Kim, M.W. Science 1993, *261*, 1018–1021.

44. Wu, X.Z.; Sirota, E.B.; Sinha, S.K.; Ocko, B.M.; Deutsch, M. Phys. Rev. Lett. 1993, *70*, 958–961.
45. Sirota, E.B. Langmuir 1998, *14*, 3133–3136.
46. Petitjean, D.; Pierre, M.; Hubert, N.; Bouroukba, M.; Dirand, M. Fuel 2001, *80*, 2007–2011.
47. Cabrera, N.; Vermilyea, D.A. *Growth and Perfection of Crystals*; Doremus, R.H., Roberts, B.W., Turnbull, J., Eds.; Wiley: New York, 1958; 393–410.
48. Davis, K.J.; Dove, P.M.; de Yoreo, J.J. Science 2000, *290*, 1134–1137.
49. Myers, D. *Surfaces, Interfaces, and Colloids*; Wiley–VCH: New York, 1999.
50. Damin, B.; Faure, A.; Denis, J.; Sillion, B.; Claudy, P.; Letoffe, J.M. Society of Automotive Engineers, Inc. 1987, Report No. 861527.
51. Lasaga, A.C. *Kinetic Theory in the Earth Sciences*; Princeton University Press: Princeton, NJ, 1998.
52. Albon, N.; Dunning, W.J. Acta Crystallogr. 1962, *15*, 474–476.
53. Boistelle, R. *Interfacial Aspects of Phase Transformations*; Mutaftschiev, B., Ed.; D. Reidel: Dordrecht, 1982.
54. Potapenko, S.Y. J. Cryst. Growth 1993, *133*, 147–154.
55. Potapenko, S.Y. J. Cryst. Growth 1993, *133*, 141–146.
56. Eidelman, N.; Azoury, R.; Sarig, S. J. Cryst. Growth 1986, *74*, 1–9.
57. Davey, R.J. J. Cryst. Growth 1976, *34*, 109–119.
58. El-Gamal, I.M. Colloid Surfaces A 1998, *135*, 283–291.
59. van Enckevort, W.J.P.; van den Berg, A.C.J.F. J. Cryst. Growth 1998, *183*, 441–455.
60. van Rosmalen, G.M.; Bennema, P. J. Cryst. Growth 1990, *99*, 1053–1060.
61. Lian, L.; Tsukamoto, K.; Sunagawa, I. J. Cryst. Growth 1990, *99*, 150–155.
62. Weissbuch, I.; Popovitz-Biro, R.; Lahav, M.; Leiserowitz, L. Acta Crystallogr. 1995, *B51*, 115–148.
63. Lahav, M.; Leiserowitz, L. J. Phys. D 1993, *26*, B22–B31.
64. Kern, R. *Growth of Crystals*; Sheftal', N.N., Ed.; Consultants Bureau: New York, 1969.
65. Koutsoukos, P.G.; Kontoyannis, C.G. J. Cryst. Growth 1984, *69*, 367–376.
66. Weijnen, M.P.C.; van Rosmalen, G.M. J. Cryst. Growth 1986, *79*, 157–168.
67. Dirksen, J.A.; Ring, T.A.; Duvall, K.N.; Jongen, N. Int. J. Refrig. 2001, *24*, 856–859.
68. Domopoulou, A.; Michaelides, A.; Skoulika, S.; Kovala-Demertizi, D. J. Cryst. Growth 1998, *191*, 166–168.
69. Englezos, P. Ind. Eng. Chem. Prod. Res. Dev. 1993, *32*, 1251–1274.
70. Kvenvolden, K.A. Rev. Geophys. 1993, *31*, 173–187.
71. Larsen, R.; Knight, C.A.; Sloan, E.D. Jr. Fluid Phase Equilibrium 1998, *150–151*, 353–360.
72. Lederhos, J.P.; Long, J.P.; Sum, A.; Christiansen, R.L.; Sloan, E.D. Jr. Chem. Eng. Sci. 1996, *51*, 1221–1229.
73. Makogon, T.Y.; Larsen, R.; Knight, C.A.; Sloan, E.D., Jr. J. Cryst. Growth 1997, *179*, 258–262.
74. Sloan, E.D.; Subramanian, S.; Matthews, P.N.; Lederhos, J.P.; Khokhar, A.A. Ind. Eng. Chem. Res. 1998, *37*, 3124–3132.

75. Amathieu, L.; Boistelle, R. J. Cryst. Growth 1986, *79*, 169–177.
76. Yeh, Y.; Feeney, R.E. Chem. Rev. 1996, *96*, 601–617.
77. DeVries, A.L. Comp. Biochem. Physiol. 1988, *90B*, 611–621.
78. Brecevic, L.; Kralj, D. J. Cryst. Growth 1986, *79*, 178–184.
79. Wesson, J.A.; Worcester, E.M.; Wiessner, J.H.; Mandel, N.S.; Kleinman, J.G. Kidney Int. 1998, *53*, 952–957.
80. Addadi, L.; Weiner, S. Angew. Chem. Int. Ed. Eng. 1992, *31*, 153–169.
81. Goldberg, H.A.; Warner, K.J.; Li, M.C.; Hunter, G.K. Connect Tissue Res. 2001, *42*, 25–37.
82. Hunter, G.K.; Hauschka, P.V.; Poole, A.R.; Rosenberg, L.C.; Goldberg, H.A. Biochem. J. 1996, *317*, 59–64.
83. Sikes, C.S.; Wheeler, A.P. Chemtech 1988, *1988*, 620–626.
84. Annual Book of ASTM Standards 5.02:D 3117-96, 2002.
85. Turnbull, D.; Cormia, R.L. J. Chem. Phys. 1961, *34*, 820–831.
86. Oliver, M.J.; Calvert, P.D. J. Cryst. Growth 1975, *30*, 343–351.
87. Kok, M.V.; Létoffé, J.-M.; Claudy, P.; Martin, D.; Garcin, M.; Volle, J-L. Fuel 1996, *75*, 787–790.
88. Claudy, P.; Létoffé, J.-M.; Neff, B.; Damin, B. Fuel 1986, *65*, 861–864.
89. Dunn, R.O. J. Am. Oil Chem. Soc. 1999, *76*, 109–115.
90. Petruccelli, S.; Añón, M.C. J. Am. Oil Chem. Soc. 1991, *68*, 684–686.
91. Davey, R.J.; Mullin, J.W.; Whiting, M.J.L. J. Cryst. Growth 1982, *58*, 304–312.
92. Roberts, K.J.; Sherwood, J.N.; Stewart, A. J. Cryst. Growth 1990, *102*, 419–426.
93. Chen, B.D.; Garside, J. J. Cryst. Growth 1996, *166*, 1094–1098.
94. van Hoof, P.J.C.M.; van Enckevort, W.J.P.; Schoutsen, M.; Bennema, P.; Liu, X.Y. J. Cryst. Growth 1998, *183*, 641–652.
95. Higgs, P.G.; Ungar, G. J. Chem. Phys. 1994, *100*, 640–648.
96. Organ, S.J.; Barham, P.J.; Hill, M.J.; Keller, A.; Morgan, R.L. J. Polym. Sci. B 1997, *35*, 1775–1791.
97. Asbach, G.I.; Kilian, H.G. Polymer 1991, *32*, 3006–3012.
98. Coutinho, J.A.P.; Ruffier-Méray, V. Ind. Eng. Chem. Res. 1997, *36*, 4977–4983.
99. Hillner, P.E.; Manne, S.; Hansma, P.K.; Gratz, A.J. Faraday Discus. 1993, *95*, 191–197.
100. Land, T.A.; Martin, T.L.; Potapenko, S.; Palmore, G.T.; de Yoreo, J.J. Nature 1999, *399*, 442–445.
101. Duffy, D.M.; Rodger, P.M. J. Phys. Chem. B 2002, *106*, 11,210–11,217.

6

Confinement and Shear Effects on the Structure of a Smectic Liquid-Crystal Complex Fluid

STEFAN H. J. IDZIAK University of Waterloo, Waterloo, Ontario, Canada

I. INTRODUCTION

The effects of confinement on complex fluids are of great technological and fundamental interest. Confined fluids are commonplace, ranging from the thin boundary lubricant between two sliding metal plates, to the interaction of micromachines with fluids, to the growth and internal operation of living cells, and to filtration.

From a more fundamental perspective, the structure of a complex fluid can be greatly altered by means of applying an external confining constraint. Both thermodynamic and rheological properties can be affected [1,2]. The liquid-crystal phase transitions can be suppressed through confinement [3], and confinement-induced solidification has also been observed [4]. The viscosity of a confined fluid generally is quite different from the bulk value [5].

For the purpose of this chapter, confinement effects can occur at many length scales, depending on the intrinsic structure found in the complex fluid. The length scales can refer to the size of an individual crystallite or oriented domain, the diameter of a colloidal particle, the radius of gyration of a polymer, or, ultimately, the size of an individual molecule or monomer. A free, or unconfined, particle is shown in Fig. 1a and the same particle confined between two rigid walls is shown in Fig. 1b. Note that confinement does not necessarily require two (or more) rigid confining surfaces; confinement can occur within a porous material or by other means. A freely suspended liquid-crystal film, for example, is confined by two air–film interfaces, and a spin-coated polymer film is confined between the rigid substrate and a single air–film interface. Polymers can also be confined between adjacent membranes in a lyotropic liquid crystal [6,7].

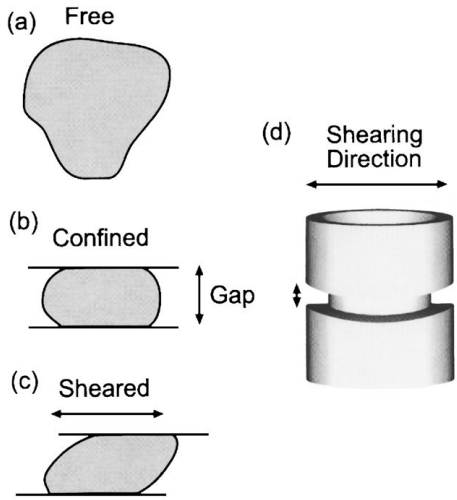


Figure 1 (a) Schematic of a free (nonconfined) drop of fluid. (b) Sketch of the fluid confined between parallel plates with gap smaller than the free-drop size. (c) Drawing of the confined fluid as it is sheared, with the upper shearing surface moving to the right. (d) Sketch of a fluid confined between two crossed cylinders, as described in the text.

Many methods are available for confining a fluid so that it can be studied in the future using a variety of probes. Fluids can be drawn into the pores of materials such as Vycor glass and aerogels. These pores have a size in the nanometer range, and once there, fluids can be held for further study [3,8–12]. Technological advances have also enabled the fabrication of microchannel devices with features in the micron to submicron range. Fluids can then be placed into these channels and examined [13,14]. Many different probes have been utilized for studying these confined fluids, including calorimetry [15], spectroscopy [2], and molecular dynamics simulations [16,17].

The above-described techniques allow the complex fluid to be studied while confined in a fixed gap. However, it is often desirable to control the degree of confinement, for example, by placing the fluid between two parallel surfaces whose separation, or gap, can be controlled, as seen in Fig. 1b. The surface forces apparatus (SFA) [18–20] allows for surface normal forces to be probed in fluids squeezed in gaps ranging from angstroms to microns. In addition, one of the surfaces can be sheared back and forth as shown in Fig. 1c, allowing for the study of viscous properties as well as other dynamic effects.

The fabrication of controllable, flat, parallel, confining surfaces with extremely small gaps can be technologically challenging. The SFA therefore

uses two crossed cylindrical lenses, as shown in Fig. 1d, to provide the confinement. The lenses are traditionally covered with atomically smooth mica. Pressure can be used to squeeze the glue holding the mica to provide a small flat region between the two mica surfaces.

Whereas calorimetric, spectroscopic, SFA, and simulation studies provide considerable information, the use of a direct structural probe such as neutron or X-ray diffraction is essential for determining the structure of the confined complex fluid. Several such studies have been completed and include the following examples: the nematic–smectic transition in the liquid crystal 4-cyano-4′octylbiphenyl (8CB) confined in an aerogel was described using X-ray diffraction (XRD) [11,12,15]; water/lutidine and polystyrene mixtures confined in the pores of a Vycor glass were examined using small angle neutron scattering (SANS) [8–10], diblock copolymer systems confined between a solid substrate and an evaporated SiO_x layer were characterized using neutron reflectivity [21,22]; adsorbed polymers confined between rigid quartz surfaces have been studied using neutron reflectivity [23–25]; and micropatterned channels and XRD have been used to characterize phase transitions in 8CB [14] and DNA alignment effects [13]. Note that most of these studies, where the separating gap can be adjusted, focus on the out-of-plane structure. The remainder of this chapter will focus on studies of the domains and in-plane structure of 8CB subjected to both confinement and shear using specially designed X-ray compatible confinement cells.

II. X-RAY CONFINEMENT CELLS

The original X-ray surface forces apparatus (XSFA) [26–33] was adapted from the MkIII surface forces apparatus (SFA) [34] designed by Israelachvili and co-workers. The SFA allowed for two mica-coated, cylindrical glass disks to be brought together and the separation carefully controlled in the region of angstroms to millimeters. The mechanism in the SFA consisted of a double cantilever spring driven by a differential spring mechanism for medium-gap control and a piezoelectric device for angstrom-level control. Surface normal forces were measured as the surface separation was changed by determining the deflection of a spring holding one of the surfaces and applying Hooke's law ($F = kx$, where F is the force, k is the spring constant, and x is the displacement). In addition, one of the surfaces can be sheared with respect to the other permitting viscosity measurements. This allows for simultaneous studies of confinement and shear.

The XSFA is based on the SFA, with the exception of the surfaces and the optics used to measure the surface separation, both of which were modified to facilitate X-ray access to the confined fluid. Figure 2a shows a schematic of the X-ray beam passing through the XSFA, scattering an angle 2θ from the

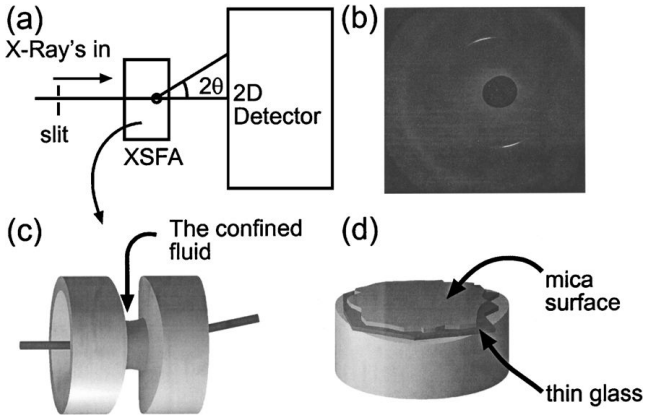


Figure 2 (a) Sketch of the incident and scattered X-ray beams passing through the (XSFA) and striking the 2D detector. (b) Typical diffraction pattern resulting from a partially oriented, confined smectic liquid crystal. (c) Schematic of the sample confined between the two confining crossed cylinders, illustrating the X-ray beam passing in transmission through the sample. (d) Drawing of the hard surfaces described in the text, showing the thin mica and quartz sheets covering the hole in the cylindrical mount. Reprinted with permission from [29]. Copyright (1996) by the American Physical Society.

confined fluid before striking a two-dimensional (2D) detector. The X-rays are collimated by a small slit placed before the XSFA as shown. A typical 2D diffraction pattern is shown in Fig. 2b, where the light regions denote areas struck by large X-ray intensities. Figure 2c shows a closeup of the confined fluid between the two cylindrical surfaces with X-rays passing through. Note that unlike the neutron reflectivity measurements described earlier, in which out-of-plane structure is determined, it is the in-plane structure that is measured with the XSFA.

The solid glass cylindrical lenses used as surface mounts in the SFA attenuate the X-rays and are therefore unsuitable for X-ray experiments. Three different types of confining surface were developed for the XSFA: capillary surfaces, hard surfaces, and soft surfaces, all of which are covered by a thin, atomically smooth mica layer in contact with the fluid. The first surfaces developed were the capillary surfaces [26–28] (not shown). These were fabricated using standard 2-mm-diameter quartz X-ray capillary tubes with 0.01 mm wall thickness and 10 mm length (the same type of capillary used for powder diffraction). A thin silver layer was evaporated on the back side of the mica sheets to form a partially reflecting mirror before gluing, silver-side down, on the capillary. Yellow light from a sodium source illuminated the

surfaces along the same axis as the X-rays. This light was partially reflected by the silver half-mirror on each surface. The resulting optical interference yielded a Newton's rings pattern from which the separation between the two mica surfaces could readily be determined.

There are two practical difficulties with capillary surfaces. The first is that considerable material (when compared with the amount of confined fluid under study) is present near the sample position, which will contribute to the measured scattering intensity. This material includes quartz, mica, and glue. The mica forms a single crystal and its contribution to the measured scattering intensity can usually be neglected. The quartz produces some background scattering that can also be accounted for. The organic glue, however, can easily have a thickness comparable to the studied confined fluid. In addition, it may have a structure with length scales comparable to the confined fluid as well, making distinguishing the glue scattering from the sample scattering quite difficult.

The second difficulty with capillary surfaces is their small radius of curvature (~1 mm). When the two cylindrical surfaces are in contact, there is only a single point of contact and the distance between the surfaces increases with radial distance from the point of contact (the separation profile between two crossed cylinders is equivalent to the distance profile between a sphere and a flat plane, to first order). The X-ray beam that illuminates the sample is typically 0.125 mm in diameter, and so, even when the surfaces are in contact, the X-ray beam can probe a sample confined in a gap as large as 1.9 μm, simply because of the curvature of the surfaces.

To compensate for both these problems, hard and soft surfaces [29,30,32,33] were developed, with an example of a hard surface shown in Fig. 2d. Aluminum disks are machined with one end having a radius of curvature of 20 mm. A 1-mm tapered hole drilled completely through the aluminum disk allows for both X-ray and optical access. The hole is then covered with a shard of very thin quartz that is then covered by a thin piece of back-silvered mica, making for a flatter, rigid surface. This results in both quartz and glue still present close to the scattering volume, but does make the range of gaps probed by the X-ray beam much smaller than with the capillary surfaces. Soft surfaces are fabricated in the same way, except that the quartz shard is omitted and the hole is covered only by the thin, back-silvered mica sheet, resulting in a soft, reasonably flexible surface with no glue or quartz near the scattering volume.

The white-light, multiple-beam-interference technique used for high-precision, small-gap measurements in the SFA was recently incorporated into the XSFA [35,36] to allow for the simultaneous measurement of surface normal forces, as well as in-plane structure using XRD. However, so far these measurements have not been made concurrently, as the sample volumes available

for scattering when the gap is tens to hundreds of angstroms wide (needed for force measurements) are not large enough for performing scattering experiments. In addition, the diffraction portion of the measurement would otherwise be extremely difficult to do, as, presumably, the fluid becomes better ordered at small gaps, behaving more like a single crystal (or singly oriented domain), which makes finding any Bragg diffraction spots quite difficult.

Recently, the X-ray confinement cell (XCC) [37,38], optimized for XRD measurements, was constructed. A photograph of the XCC is shown in Fig. 3. This instrument still uses the crossed cylinder surfaces of the XSFA but has been optimized for diffraction experiments by improving long-time stability

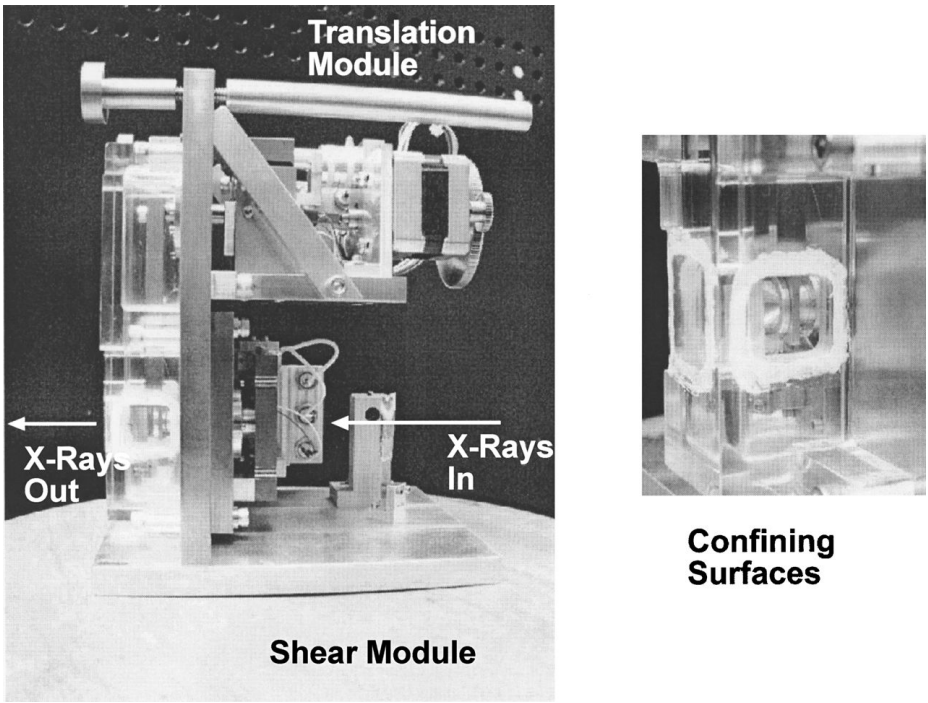


Figure 3 Photograph of the X-ray confinement cell showing the path of the X-ray beam passing through the confining surfaces (magnified at right). The translation module at top and the shear module at the bottom can be seen. The Plexiglas cover is used to both hermetically seal the sample environment and to keep drafts off the sample, which can affect the stability of the confining gap.

by means of capacitor feedback translation stages and the use of high-precision optical encoders. In addition, position feedback has been added to the shearing mechanism, making it much stiffer, which facilitates the study of high-viscosity complex fluids. Although this instrument is not well suited to performing force measurements at very close separations, it is ideally suited for the study of fluids confined in gaps ranging from ~500 Å to millimeters.

III. EXPERIMENTAL DETAILS

The liquid-crystal 8CB, which can form crystalline, nematic, and smectic phases depending on temperature, is pictured in Fig. 4a. Two 8CB molecules are arranged together to form a dimer. At room temperature, these long rodlike dimers arrange themselves into planes, as shown in Fig. 4b, forming a smectic phase. There is liquidlike order between the dimers within a single plane, while the planes are packed in a regular fashion, with a nominal spacing of 31.73 Å. The smectic phase therefore takes on both the structural characteristics of a disordered liquid and an ordered crystal, hence a liquid crystal. 8CB is an ideal candidate for studies of a confined fluid. It is extremely well characterized in the bulk under both static and shear conditions, and some studies of 8CB confined in porous materials have been completed [11,15]. In addition, its structural properties under shear have been carefully studied using a Couette shear cell and synchrotron X-ray diffraction [40–42].

Simple shear in a fluid occurs when the fluid is placed between two plates, one of which is moving, as shown in Fig. 4c. If the fluid is Newtonian, nonslip boundary conditions imply a linear velocity gradient in the fluid between the two walls. The shear rate, $\dot{\gamma} = v/s$, is given by the velocity (v) divided by the gap between the two walls (s). In a Couette shear cell, this shear rate is produced using the gap between two concentric cylinders, one of which is rotating, whereas in the XSFA and XCC, this is accomplished by oscillating one surface back and forth with respect to the other.

Using the notation of Miecowicz [39], there are three orthogonal orientations that a layered material, such as a smectic liquid crystal can take under shear, as shown in Fig. 4d. The orientation of the smectic layers can be described by the layer normal, **n**, which is a unit vector perpendicular to the layers themselves. The shear profile can be described using two directions, the velocity direction of the flow itself, **v**, and the velocity gradient direction, $\nabla \mathbf{v}$, which is perpendicular to the two shearing surfaces. In the "a" orientation, the smectic layer normal is orthogonal to both the velocity and velocity gradient directions. The layer normal is parallel to the velocity direction in the "b" orientation, whereas the layer normal is parallel to the velocity gradient direction in the "c" orientation. Intuitively, one would expect that both the

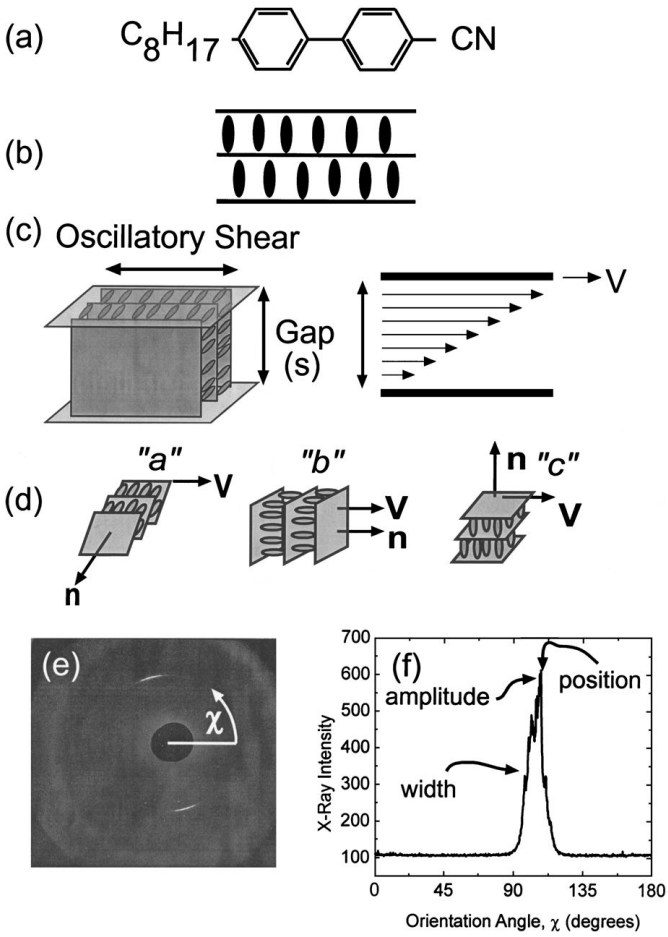


Figure 4 (a) The liquid crystal 8CB forms a layered smectic phase at room temperature, as shown in (b). (c) Sketch of one of the possible orientations of the smectic layers while applying an oscillating shear to the upper confining surface. The oscillating surface produces a velocity gradient between the surfaces as shown. (d) Schematic of the three orthogonal orientations of a smectic phase under shear, using the notation of Miecowicz [39]. (e) Typical 2D X-ray diffraction pattern of 8CB under confinement. Note that the existence, or lack thereof, of the "c" orientation cannot be determined in these experiments using the X-ray confinement cell. (f) Orientation scan derived from the 2D image showing the peak position, amplitude, and width.

"a" and "c" orientations would be possible in a sheared liquid crystal, but that the "b" would be intrinsically unstable, as the shear would tend to deform the smectic layers. In fact, the "b" orientation was never observed in the synchrotron XRD studies of the bulk 8CB sheared in a Couette cell whereas both the "a" and "c" orientations were observed, depending on temperature and shear rate.

A typical 2D XRD pattern of 8CB, confined between two hard surfaces in a gap of 5 μm, is shown in Fig. 4e. The dark disk at the center is the shadow of the beam stop, which protects the detector from exposure to the direct, incident beam. The scattering from the smectic layers can be seen in the two light-colored arcs. Considerable information can be gathered from this image. The radial distance of the diffraction peaks from the center can give the *d*-spacing between the smectic layers by means of Bragg's law, $\lambda = 2d \sin \theta$, where λ is the X-ray wavelength, d is the interlayer spacing, and θ is half the scattering angle. The *d*-spacing of 8CB remains constant for all the measurements discussed in the upcoming sections and never deviates from its bulk value.

The orientation or mosaic scan shown in Fig. 4f is derived from measuring the scattering intensity as a function of the orientation angle, χ, as shown in Fig. 4e. The plot in Fig. 4f describes a reasonably well-oriented sample. Note that a completely unoriented or randomly oriented sample would produce a uniform circle of scattering in the 2D image and a flat line in the orientation scan. The peak position describes the mean angle at which the smectic layer normals are pointing—in this case ~110°. The peak width, $\Delta\chi$, describes the degree of orientation, or mosaic, of the sample. The sharper the peaks, the better oriented the sample. The peak amplitude is loosely related to the amount of oriented material present in the sample, assuming that the mosaic does not change.

IV. COMPETITION BETWEEN SHEAR AND CONFINEMENT

Capillary surfaces were used in the first studies of 8CB under confinement in the XSFA [26–28]. A schematic of the two surfaces and the direction of shear of the top surface is shown in Fig. 5a. This unusual shear direction (future studies were performed with the shear parallel to the axis of the shearing surface) was used in an attempt to probe the "c" orientation. This orientation cannot ordinarily be probed while passing the X-ray beam in transmission through the sample. Instead, the X-ray beam was passed between the two surfaces, to probe the "c" orientation in reflection. The "c" orientation was not, in fact, detected, which is consistent with previous work which showed

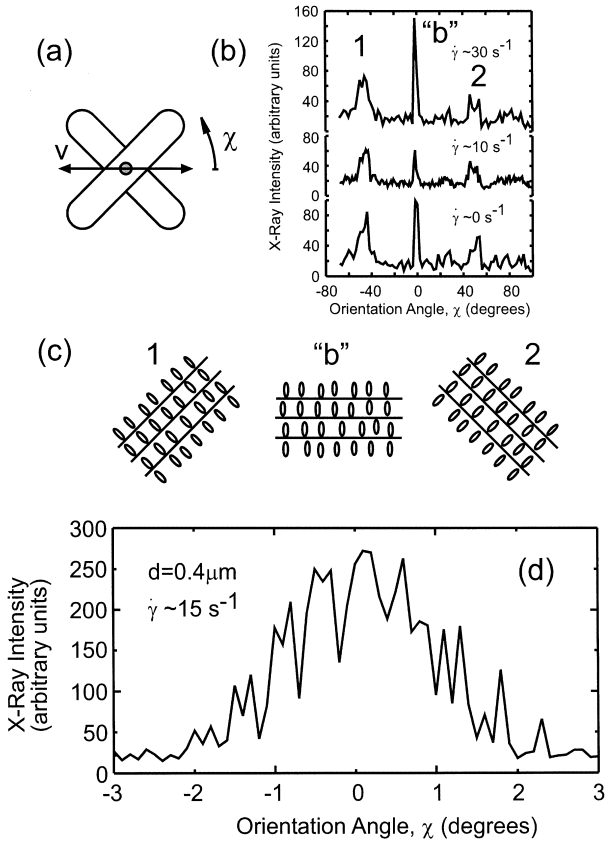


Figure 5 (a) Drawing of the two crossed cylindrical surfaces showing the shearing direction of the upper surface. (b) Orientation scans for 8CB confined in a 0.4-μm gap between capillary surfaces for shear rates of 0, 10, and 30 s^{-1}. (c) Representation of the three orientations of smectic layers corresponding to the three peaks seen in (b). (d) High-resolution mosaic scan corresponding to the "b" orientation, clearly showing the presence of discrete domains. Reprinted with permission from [26]. Copyright (1994) American Association for the Advancement of Science.

that 8CB molecules lie in a planar conformation on mica, with the molecules lying flat on the surface [43].

The gap between the two mica surfaces was set to ~0.4 μm and orientation scans were performed at various shear rates, as seen in Fig. 5b. Note that, unlike later studies using the XSFA, a 2D detector was not used in this study. Instead, a traditional point detector and four-circle goniometer were

employed to determine the orientation by keeping the position of the detector fixed and rotating the XSFA. This meant that data acquisition was very slow—on the order of hours per measurement.

In the case of no applied shear, three peaks in the orientation scan can be clearly seen at 0° and ± 45°. These peaks correspond to the three orientations sketched in Fig. 5c, where the orientations labeled "1" and "2" correspond to smectic layers oriented with layer normals parallel to the axis of each crossed cylindrical surface. The third peak corresponds to the "b" orientation. Note that it is not obviously clear why any orientation occurs at zero shear; possible mechanisms for the orientation will be discussed later.

Of considerable interest are the orientation scans shown in Fig. 5b for shear rates of 10 and 30 s^{-1}. There is essentially no difference between the case of zero shear and applied shear. The "1" and "2" orientations remain, and, more importantly, the "b" orientation also recurs. This latter orientation was never seen in the bulk flow studies, even at very low shear rates, which implies that confinement-induced orientational effects are dominating over shear alignment. Another effect that occurs can be seen in Fig. 5d, which is a higher-resolution orientation scan of the "b" peak. It can clearly be seen that this is not the expected smooth peak. Instead, it consist of a series of many discrete peaks, implying that several discrete, singly oriented domains of the smectic liquid crystal are present confined between the mica surfaces.

These phenomena suggest that the alignment imposed by confinement is quite strong, preventing the usual shear-induced orientation that occurs in bulk shear flow studies of 8CB. In addition, the small sample volume probed by the X-ray beam contains only a small number of discrete liquid-crystalline domains, which is manifested by the discrete peaks in the high-resolution orientation scan.

The original alignment of the material can be due to several factors. As stated earlier, 8CB takes on a planar conformation on a mica surface. Therefore, the initial orientation would be expected to lie in the "a"–"b" plane, with any "c" orientation excluded. Considerable bulklike flow occurs during sample loading and during the time the surfaces are being brought together, expelling material from between the surfaces. The capillary surfaces are quite curved, with the radius of curvature ~1 mm. An orientation with smectic-layer normals parallel to the cylindrical axis of the surface (i.e., with layers oriented like disks radially around the cylinder) would not contain any defects due to the curvature of the surfaces. This orientation corresponds to the "1" and "2" peaks in the orientation scans. Other orientations would require the formation of defects in the smectic layers to grow and are therefore energetically unfavorable. The observed "b" orientation presumably arises from the strong flow occurring during sample loading and is then pinned in place by the confining surfaces. Interestingly, the overall width of the high-

resolution "b" peak (~3.6°) roughly corresponds to the angular deviation the cylindrical surface makes from a flat plane over the 0.125-mm size of the X-ray beam probing the sample.

V. SHEAR-INDUCED ORIENTATION

The liquid crystal 8CB was studied [29,30,32,33] under shear using the previously described hard surfaces with the much larger (20 mm) radius of curvature. Two-dimensional images of the diffraction patterns observed in the case of zero applied shear and 125 s^{-1} applied shear are shown in Figs. 6a and 6b. The initial relatively poorly aligned orientation is quite different from that

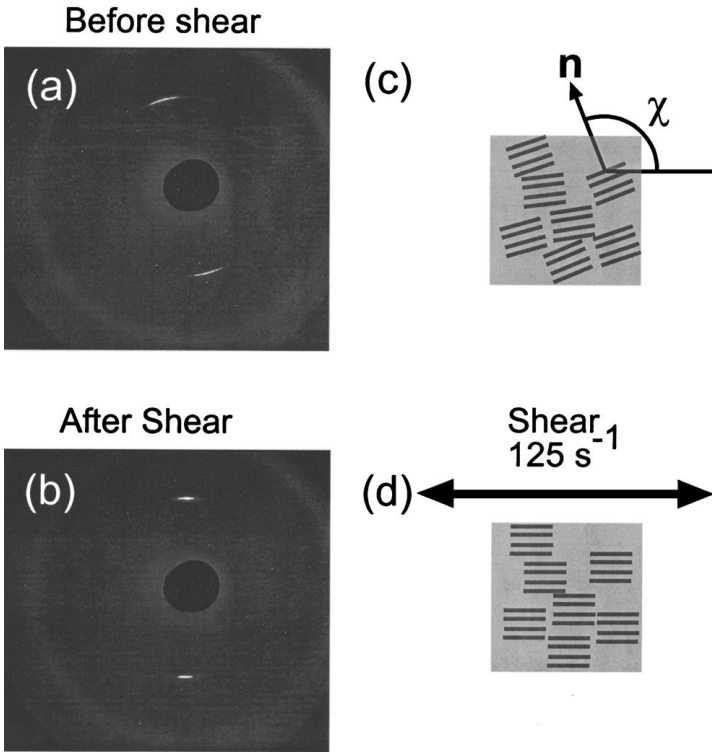


Figure 6 (a) Typical 2D diffraction pattern of 8CB confined between hard surfaces with a gap of 5 μm before shear and (b) while a shear rate of 125 s^{-1} is applied. (c) Schematic representation of the smectic layers before shear. χ is the angle describing the direction of the plane normal. (d) Sketch of the orientation of shear-oriented smectic domains.

seen for capillary surfaces and, in fact, only a single orientation peak can be seen, corresponding to the smectic domains oriented as illustrated in Fig. 6c. This initial orientation is presumably mostly affected by the bulklike flow conditions that occur during sample loading. The effects of surface curvature would not be expected to play as large a role as with the very curved capillary surfaces, hence the lack of orientation in the "1" and "2" directions. The orientation changes quite dramatically after the application of shear, with the diffraction peaks both narrowing and rotating, implying a better oriented material with domains rotated from the zero shear case as sketched in Fig. 6d.

This shear-induced orientation corresponds to the "a" orientation, which has been observed in bulk shear studies. The degree of orientation achievable with shear in the XSFA is quite impressive, with mosaics on the order of 2° achievable. In addition to 8CB, the XSFA has been used to orient otherwise difficult to orient zwitterionic copolymers [32,33] and hexagonal [30,33] lyotropic liquid crystals and has also been used to orient lamellar [36] lyotropic liquid crystals. This ability to orient otherwise difficult to align materials could be quite useful, as oriented samples are needed to probe very weak diffraction peaks and are also needed to properly study very weak thermal diffuse scattering, which can be used to examine the elastic moduli. Presumably, this orientation technique could also be used to produce highly oriented films on substrates as well, resulting in coatings with a variety of engineered properties.

VI. MECHANICAL PROPERTIES OF CONFINING SURFACES

The effects of the mechanical properties of the confining surfaces on the nature of confinement can be probed using both hard and soft surfaces [29]. Both of these surfaces are geometrically similar, having an identical radius of curvature of 20 mm, much greater than capillary surfaces. However, the mechanical rigidity of the two types of surface is quite different. In the case of hard surfaces, the inherent elasticity of the confined 8CB is not strong enough to deform the surface geometry. This can be easily seen by observing the Newton's rings interference pattern that results from shining monochromatic sodium light through the two back-silvered mica surfaces. The resulting interference pattern is a very sensitive probe of the separation between the two surfaces as a function of position. The Newton's rings pattern does not change at all and remains circular for 8CB confined in a variety of gaps, from submicron to tens of microns.

The rigidity of the soft surfaces, consisting only of a several-micron-thick sheet of mica covering a 1-mm hole in the cylindrical surface mount, is quite

low. These surfaces are easily deformable when 8CB is confined in small gaps on the order of several microns, where the sample elasticity competes quite favorably with the surface rigidity. This is confirmed by the large deformations observed in the Newton's rings pattern.

The use of a 2D image plate X-ray detector decreases the data acquisition time to ~20 min per experiment, in contrast to the several hours required with the point detector. This facilitates systematic structural studies of materials under a variety of gaps, as well as shear rates.

Orientation scans for 8CB confined between two soft surfaces under several gaps are shown in Fig. 7a. Note that the scale is magnified for the 1-μm-gap case. As can clearly be seen, the width of the peak, which represents the mosaic, or degree of orientational ordering, is relatively large at large gaps, decreases at intermediate gaps, and increases again at smaller gaps. Similar orientation scans for 8CB confined between hard surfaces are shown in Fig. 7b, where it is observed that the peak width remains constant with changing gap.

The peak widths derived from both hard- and soft-surface orientation scans plotted as a function of confining gap are plotted in Fig. 7c. The mosaic is essentially constant for hard-surface confinement, decreasing only slightly as the gap increases. Of great interest is the observation that the mosaic decreases to a minimum value at a critical gap of ~3.4 μm for soft-surface confinement, before increasing again at larger gaps. Both types of surface have atomically smooth mica in contact with the 8CB, the only difference is the mechanical rigidity of the confining surfaces. Only the flexibility of the surfaces can be responsible for the difference in the orientational ordering of the fluid. Note that this behavior is reversible and occurs as the gap is increased or decreased.

Schematics of the 8CB domains confined between soft surfaces are shown for small, critical, and large gaps in Figs. 7d–7f, respectively. This coarsening of the orientational order can be understood by considering what occurs as the gap is increased. Presumably, some form of epitaxial alignment pins the orientation of the domains in contact with the mica surfaces themselves. As the cylinders are not exactly crossed 90° with respect to each other and the crystalline lattices of the two mica surfaces are not precisely aligned, it would be expected that the orientation of the 8CB layers making contact with the top surface would be slightly different than those making contact with the bottom surface.

At very small gaps, the domains span both surfaces, so competition between the two epitaxial-induced orientations will occur, resulting in the relatively broad mosaic seen at small gaps. Consider a domain oriented by the top surface as the gap grows. There is a defect energy associated with the domain's contact with the bottom surface, as well as a grain boundary energy associated

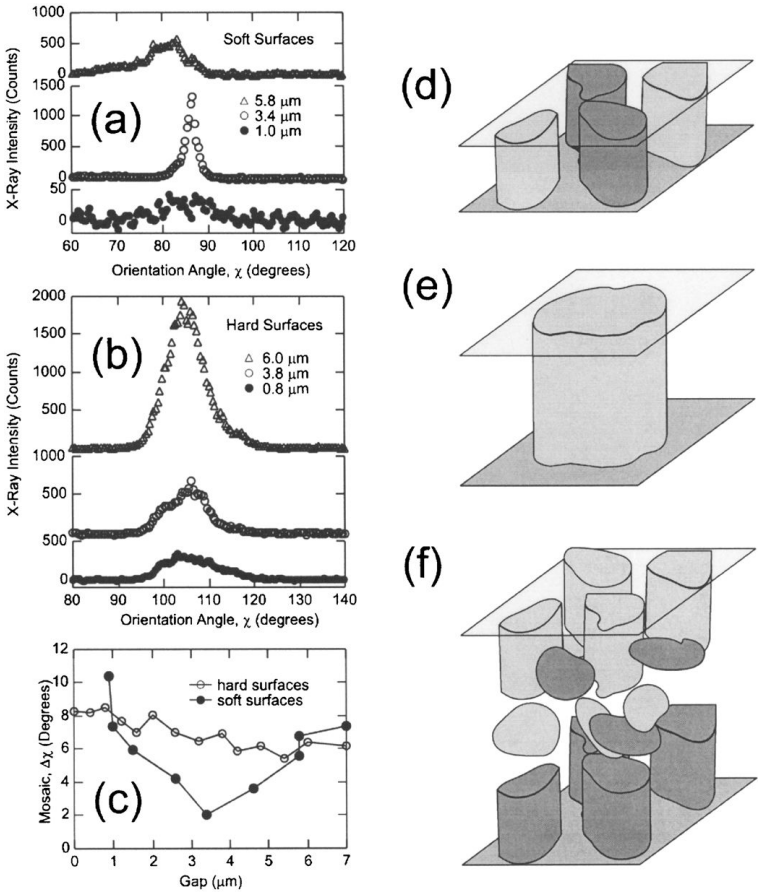


Figure 7 (a) Orientation scans for 8CB confined between soft surfaces for confining gaps of 5.8, 3.4, and 1.0 μm. (b) Orientation scans for 8CB confined between two hard surfaces for gaps of 6.0, 3.8, and 0.8 μm. (c) Plot of mosaic ($\Delta\chi$) as a function of confining gap for hard surfaces (open circles) and soft surfaces (solid circles) clearly showing the critical gap of 3.4 μm for orientational coarsening in soft surfaces. Schematic of smectic domains at small gap (d), critical gap (e), and large gap (f) for soft surfaces, showing the influence of the surfaces on the orientational order. Reprinted with permission from [29]. Copyright (1996) by the American Physical Society.

with the side wall of the domain. As the gap is increased, this grain-boundary energy increases and eventually becomes large enough that it becomes energetically favorable for several individual domains to reorient and join together to form a large singly oriented domain with smaller grain-boundary energy. This coalescing of the domains is then manifested by the appearance of a much better oriented sample with small mosaic. In the case of soft surfaces, this can occur because the surface deformability allows the defects between the domains to migrate easily, facilitating the domain fusion. The defects are pinned by the hard-surface confinement, which prevents the domains from moving and joining, resulting in no change in the observed orientational mosaics. As the surface gap is further increased, more and more material flows between the surfaces and the sample becomes more disordered, resulting in the increase in sample mosaic.

This domain-orientational coarsening is a stunning example of a structural behavior that results under confinement due to competition between the sample and confining surface elasticity. The microscopic coarsening presumably will also affect the rheological properties of the fluid, suggesting that the design of high-performance lubrication is not only affected simply by the fluid itself, or the roughness of the surface, or the chemical interaction between fluid and surface, but that the mechanical properties of the surface must also be considered.

VII. TIME-DEPENDENT STRUCTURAL EFFECTS

The advent of extremely high-brilliance third-generation X-ray synchrotron sources has allowed the use of the XCC [37], which is the successor to the XSFA, to be used to conduct time-dependent orientational studies of 8CB confined between hard mica surfaces [37,38]. The greatly increased 2D X-ray intensity incident on the sample coupled with technological improvements to 2D X-ray detectors greatly reduces data acquisition times. Good quality datasets can be acquired in less than 1 min, allowing for relatively fast phenomena to be probed.

Figures 8a–8c show the evolution of the orientational peak position, mosaic (or peak width), and peak amplitude for 8CB, confined in a relatively large gap of 5 μm, as a function of time for two very different shear rates. The difference in behavior between the two shear rates is striking. At the high shear rate, the peak position rapidly shifts to 90° as the smectic layers rotate into the "a" orientation. Similarly, the mosaic decreases rapidly and the peak amplitude increases reasonably quickly with time. At the lower shear rate, the behavior is considerably slower. The smectic-layer orientation rotates slowly but has not reached the "a" orientation after 1200 s. The mosaic is not as low as with high shear and the peak amplitude does not rise continuously,

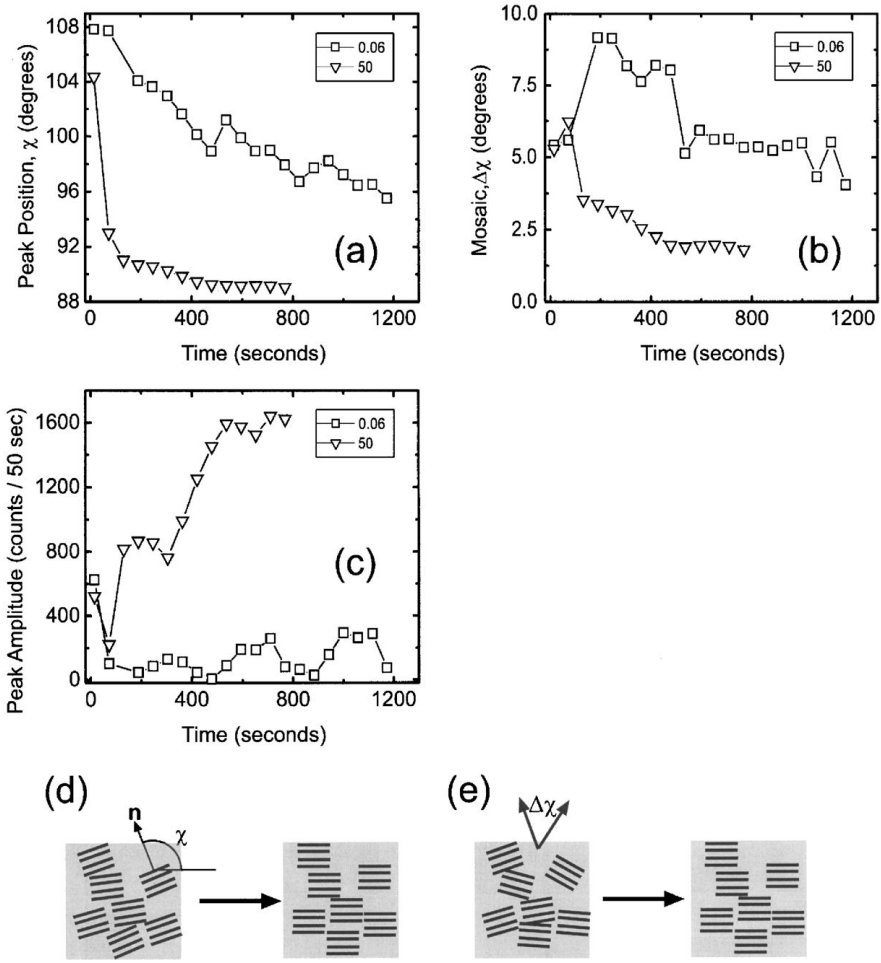


Figure 8 Orientation peak position (a), mosaic (b), and peak amplitude (c) scans as a function of time for 8CB confined between hard surfaces separated by 5 μm for two shear rates, 0.06 (squares) and 50 s^{-1} (triangles). (d) Sketch of the orientation angle of the domains initially and after shear has been applied for some time. (e) Schematic demonstrating the dramatic decrease in mosaic ($\Delta\chi$) due to shear-induced orientational ordering over time.

but actually oscillates somewhat, as will be discussed later. The behavior in general can be summarized by the sketches in Figs. 8d and 8e, where the rotation of different smectic domains over time, as well as the increased orientational order, is shown.

The behavior for 8CB confined in a 0.5-μm gap is shown in Fig. 9, where the time evolution of the peak position and mosaic are shown in Fig. 9a and 9b, respectively for two different shear rates. At the higher shear rate, the behavior is qualitatively similar to the large gap case, but slower. The smectic layers rotate into the "a" orientation, and the mosaic decreases over time, as seen in Fig. 9c, although now the mosaic changes in discrete steps, as small (or individual) groups of domains reorient at once. This behavior is shown

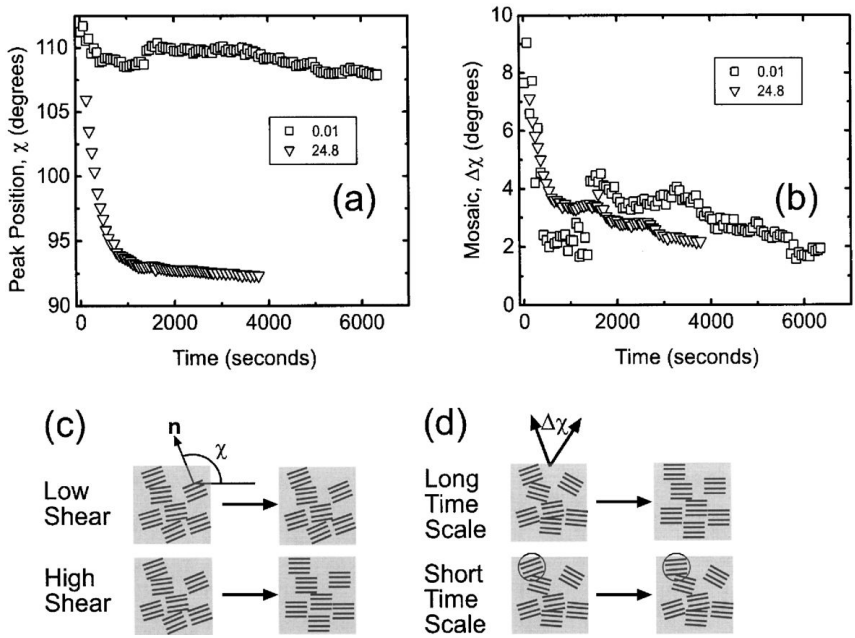


Figure 9 Plot of peak position (a) and mosaic (b) as a function of time for 8CB confined between hard surfaces separated by 0.5 μm for shear rates of 0.01 (squares) and 24.8 s^{-1} (triangles). Note the steps seen in the mosaic plots and the slight change in peak position at low shear rate. (c) Sketch of the change in smectic-layer orientation for low shear and high shear, illustrating the lack of reorientation at low shear that is present at high shear. (d) Sketch of the smectic domains illustrating the change observed in sample mosaic over short and long time periods for high shear, showing the discrete reorientation occurring in isolated domains, represented by the steps in the mosaic scan (b).

schematically in Fig. 9d. At low shear, the peak position does not change significantly with time and remains roughly constant, whereas the mosaic does change quite significantly, implying that the degree of orientational ordering improves over time, although the shear is not strong enough to rotate the mean orientation of the smectic layers.

Extremely interesting behavior can be seen in the peak amplitude plots seen in Fig. 10a for 8CB confined in a 0.5-μm gap and sheared with a rate 0.01 s^{-1}. Normally, two diffraction spots can be seen on the 2D detector. This results because the mosaic of the 8CB domains is sufficiently large that scattering can occur both in the upper and lower directions. However, in this case, where the applied shear is 0.01 s^{-1} (corresponding to a surface oscillation frequency of one back-and-forth motion every 100 s), this does not occur. This can be seen in Fig. 10b where two consecutive 2D images, taken 57 s apart, show strong scattering in the upper peak in one image, followed by strong scattering in the lower peak 57 s later. Figure 10a shows this behavior more dramatically, where the difference in amplitude between the upper and lower peaks is plotted as a function of time. Note the oscillation in intensity, which matches quite well with the 100-s period of the surface oscillation.

This phenomenon can be understood by considering the sketches shown in Figs. 10c–10g. Figure 10c shows the smectic layers oriented with layer normal ~110° from the x axis (the "a" orientation results when the smectic layer normals point along the y axis, 90° from the x axis). The upper surface oscillates back and forth parallel to the x axis and the x ray beam is incident along the z axis. Figures 10d and 10e show the deformed smectic domain as the surface is sheared to the left and then to the right, respectively. As the shearing surface moves to the left, the smectic layers are tilted slightly in the y–z plane, as seen in Fig. 10f, due to the fact that the layers are not exactly parallel to the x axis. Similarly, as the shearing surface moves to the right, 50 s later, the smectic layers tilt in the other direction, as shown in Fig. 10g. This tilting is of sufficient magnitude, and the sample mosaic is sufficiently small, that, in one case, Bragg's law cannot be satisfied for X-rays scattering up, and 50 s later, Bragg's law cannot be satisfied for X-rays scattering down. This depends on which direction the shearing surface is moving, resulting in the observed oscillation of scattering intensity between upper and lower diffraction spots. This behavior is not observed at higher surface oscillation frequencies (corresponding to higher shear rates), as the 50-s 2D detector exposure times effectively integrate out the behavior. This phenomenon is also presumably responsible for the oscillation in intensity seen in Fig. 8c.

The applied low shear at small confining gaps provides enough energy to both elastically deform the smectic domains and to improve the alignment of these domains by encouraging defects to migrate and allowing for domains to

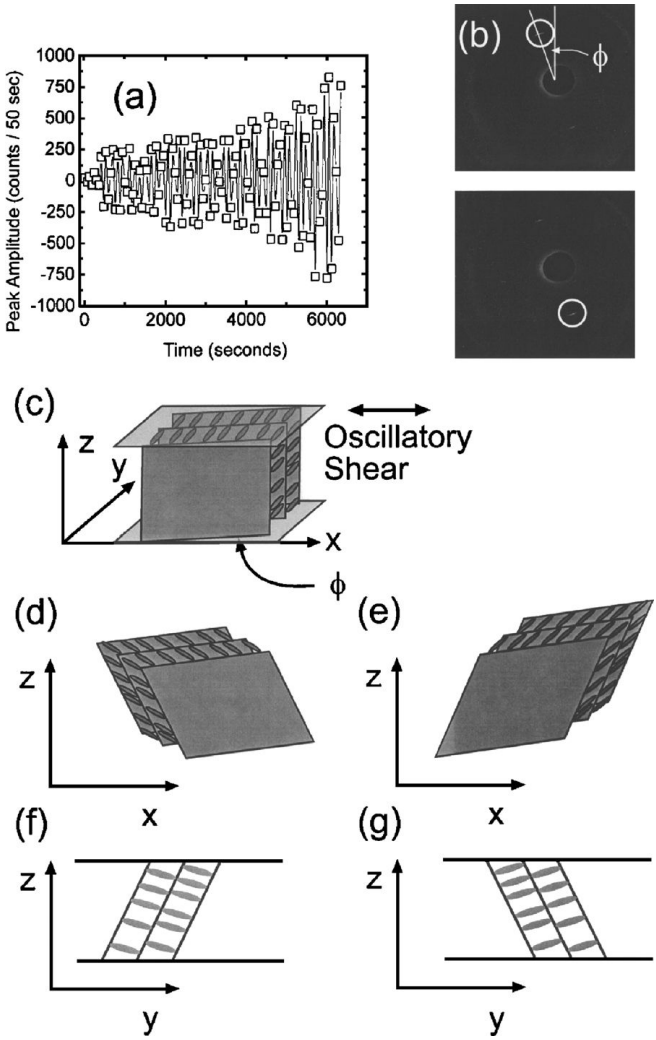


Figure 10 Plot of difference of upper and lower peak amplitudes as a function of time for 8CB confined between hard surfaces separated by 0.5 μm with applied shear of 0.01 s^{-1}. Note the oscillation in amplitude. (b) Two consecutive 2D images taken 57 s apart, clearly showing the oscillation in diffraction intensity between upper and lower diffraction peaks. (c) Schematic of the smectic layers with layer normal oriented an angle χ to the x axis, or with layers oriented an angle $\phi = \chi - 90$ to the x axis. (d) Sketch of a deformed domain as the shearing surface moves left. Because the smectic layers are not parallel to the x axis, this results in a tilt of the layers in the y–z plane (f). (e) Drawing of the smectic domain deformed as the shearing surface moves right, with corresponding tilt in the y–z plane (g).

anneal and fuse. However, the shear does not provide enough energy to rotate the collective average orientation of the domains or to alter the *d*-spacing of the smectic layers as they are deformed. At high shear rates, sufficient energy can be provided to both improve the mosaic of the smectic phase and produce a collective rotation into the "a" orientation.

VIII. CONCLUSION

There are few direct structural studies of confined fluids, primarily due to the high degree of difficulty in performing the measurements. Technological improvements in both synchrotrons and detectors have greatly facilitated the studies, although the application of both confinement and shear remains a challenge.

The studies of confined 8CB to date have been extremely fruitful, providing several interesting results. At low gap and high surface curvature, competition between confinement-induced orientation and shear-induced orientation favors confinement. However, for less curved surfaces and reasonable shear rates, the effects of shear will dominate and sample orientation can be dramatic.

The nature of the mechanical properties of confining surfaces needs to be considered to properly determine the final structural behavior of a confined fluid, which will presumably also affect technologically important rheological properties.

Time-dependent measurements have demonstrated that the discrete smectic domains can deform elastically under shear, as the shearing surface deforms the domain back and forth. At low shear rates, this elastic deformation can occur and is not strong enough to rotate the average orientation of the smectic layers; however, sufficient energy is added to anneal the smectic domains and produce a well-ordered material.

These studies suggest that the study of confined fluids will continue to be of great interest and will spawn many new results. The study of confinement on the effects of processing, (e.g., on crystal growth kinetics and biological activities occurring within cells) should prove to be fruitful.

REFERENCES

1. Drake, J.M.; Klafter, J. Physics Today 1990, *43*, 46.
2. Drake, J.M.; Klafter, J.; Levitz, P. Science 1991, *251*, 1574–1579.
3. Aliev, F.M.; Plechakov, I.V. In *Dynamics in Small Confining Systems III*; Drake, J.M. Klafter, J., Kopelman, R., Eds.; Materials Research Society: Pittsburgh, PA, 1997; Vol. 464, 177–182.

4. Klein, J.; Kumacheva, E. Science 1995, *269*, 816–819.
5. Klein, J.; Kumacheva, E.; Mahalu, D.; Perahia, D.; Fetters, L.J. Nature 1994, *370*, 634–636.
6. Radlinska, E.Z.; Gulik-Krzywicki, T.; Lafuma, F.; Langevin, D.; Urbach, W.; Williams, C.E.; Ober, R. Phys. Rev. Lett. 1995, *74*, 4237–4240.
7. Ligoure, C.; Bouglet, G.; Porte, G.; Diat, O. J. Phys. II 1997, *7*, 473–491.
8. Lin, M.Y.; Sinha, S.K.; Drake, J.M.; Thiyagarajan, P.; Wu, X.L.; Stanley, H.B. J. Phys. IV 1993, *3*, 109–115.
9. Lin, M.Y.; Sinha, S.K.; Drake, J.M.; Wu, X.-l.; Thiyagarajan, P.; Stanley, H.B. Phys. Rev. Lett. 1994, *72*, 2207–2210.
10. Lal, J.; Sinha, S.K.; Auvray, L. J. Phys. II 1997, *7*, 1597–1615.
11. Clark, N.A.; Bellini, T.; Malzbender, R.M.; Thomas, B.N.; Rappaport, A.G.; Muzny, C.D.; Schaefer, D.W.; Hrubesh, L. Phys. Rev. Lett. 1993, *71*, 3505–3508.
12. Bellini, T.; Rappaport, A.G.; Clark, N.A.; Thomas, B.N. Phys. Rev. Lett. 1996, *77*, 2507–2510.
13. Li, Y.; Wong, G.C.L.; Caine, E.; Hu, E.L.; Safinya, C.R. Int. J. Thermophys. 1998, *9*, 1165–1174.
14. Shi, Y.S.; MartinezMiranda, L.J.; Kumar, S. Mol. Cryst. Liq. Cryst. 1997, *301*, 73–78.
15. Bellini, T.; Clark, N.A.; Muzny, C.D.; Wu, L.; Garland, C.W.; Schaefer, D.W.; Oliver, B.J. Phys. Rev. Lett. 1992, *69*, 788–791.
16. Rhykerd, C.L. Jr.; Schoen, M.; Diestler, D.J.; Cushman, J.H. Nature 1987, *330*, 461–463.
17. Thompson, P.A.; Grest, G.S.; Robbins, M.A. Phys. Rev. Lett. 1992, *68*, 3448–3451.
18. Israelachvili, J.N.; McGuiggan, P.M.; Homola, A.M. Science 1988, *240*, 189–191.
19. Klein, J.; Perahia, D.; Warburg, S. Nature 1991, *352*, 143–145.
20. Van Alsten, J.; Granick, S. Phys. Rev. Lett. 1988, *61*, 2570–2573.
21. Lambooy, P.; Russell, T.P.; Kellogg, G.J.; Mayes, A.M.; Gallagher, P.D.; Satija, S.K. Phys. Rev. Lett. 1994, *72*, 2899–2902.
22. Lambooy, P.; Salem, J.R.; Russell, T.P. Thin Solid Films 1994, *252*, 75–77.
23. Cosgrove, T.; Zarbakhsh, A.; Luckham, P.F.; Hair, M.L.; Webster, J.R.P.W. Faraday Discuss. 1994, *98*, 189–201.
24. Cosgrove, T.; Luckham, P.F.; Richardson, R.M.; Webster, J.R.P.; Zarbakhsh, A. Colloid Surfaces A 1994, *86*, 103–110.
25. Kuhl, T.L.; Smith, G.S.; Israelachvili, J.N.; Majewski, J.; Hamilton, W. Rev. Sci. Instrum. 2001, *72*, 1715–1720.
26. Idziak, S.H.J.; Safinya, C.R.; Hill, R.S.; Kraiser, K.E.; Ruths, M.; Warriner, H.E.; Steinberg, S.; Liang, K.S.; Israelachvili, J.N. Science 1994, *264*, 1915–1918.
27. Idziak, S.H.J.; Safinya, C.R.; Sirota, E.B.; Bruinsma, R.F.; Liang, K.S.; Israelachvili, J.N. In *Structure and Flow in Surfactant Solutions*; Herb, C.A. Prud'homme, R.K., Eds.; American Chemical Society: Washington DC, 1994; Vol. 578, 288–299.

28. Idziak, S.H.J.; Koltover, I.; Liang, K.S.; Israelachvili, J.N.; Safinya, C.R. Int. J. Thermophys. 1995, *16*, 299–307.
29. Idziak, S.H.J.; Koltover, I.; Israelachvili, J.N.; Safinya, C.R. Phys. Rev. Lett. 1996, *76*, 1477–1480.
30. Idziak, S.H.J.; Koltover, I.; Davidson, P.; Ruths, M.; Li, Y.L.; Israelachvili, J.N.; Safinya, C.R. Physica B 1996, *221*, 289–295.
31. Idziak, S.H.J.; Li, Y. Curr. Opin. Colloid Interf. Sci. 1998, *3*, 293–298.
32. Koltover, I.; Idziak, S.H.J.; Safinya, C.R.; Steinberg, S.; Israelachvili, J.N.; Liang, K.S. Mater. Res. Soc. Symp. Proc. 1995, *366*, 101–112.
33. Koltover, I.; Idziak, S.H.J.; Davidson, P.; Li, Y.; Safinya, C.R.; Ruths, M.; Steinberg, S.; Israelachvili, J.N. J. Phys. II 1996, *6*, 893–907.
34. Israelachvili, J.N.; McGuiggan, P.M. J. Mater. Res. 1990, *5* (10), 2223.
35. Golan, Y.; Seitz, M.; Luo, C.; Martin-Herranz, A.; Yasa, M.; Li, Y.L.; Safinya, C.R.; Israelachvili, J. Rev. Sci. Instrum. 2002, *73*, 2486.
36. Golan, Y.; Martin-Herranz, A.; Li, Y.; Safinya, C.R.; Israelachvili, J. Phys. Rev. Lett. 2001, *86*, 1263.
37. Nieman, B., Commeinhes, X., Babcock, N.S., Frola, I., Forgett, R., Idziak, S.H.J., Sutton, M. Rev. Sci. Instrum. *in press*.
38. Nieman, B., Sutton, D.M., Idziak, S.H.J. *submitted*.
39. Miecowicz, M. Nature 1946, *158*, 27.
40. Safinya, C.R.; Sirota, E.B.; Bruinsma, R.F.; Jeppesen, C.; Plano, R.J.; Wenzel, L.J. Science 1993, *261*, 588–591.
41. Safinya, C.R.; Sirota, E.B.; Plano, R.J. Phys. Rev. Lett. 1991, *66*, 1986–1989.
42. Safinya, C.R.; Sirota, E.B.; Plano, R.; Bruinsma, R.F. J. Phys. Condensed Matter 1990, *2*, SA365–SA371.
43. Pieranski, P.; Jerome, B. Phys. Rev. A 1989, *40*, 317–322.

7

Macroscopic Rheological Behavior of Dispersions of Soft Rubberlike Solid Particles

RAJINDER PAL University of Waterloo, Waterloo, Ontario, Canada

I. INTRODUCTION

Dispersions of solid particles in liquids, also referred to as suspensions, are important practically. They find application in a wide variety of industries; examples of industries where dispersions are important include foods, paints, dyestuffs, pigments, paper coatings, printing, cosmetics, detergents, ceramics, pharmaceutical, pesticides, explosives, petroleum, and polymers [1–7]. Both aqueous and nonaqueous dispersions are encountered in industry.

In the formulation, handling, storage, processing, and transport of dispersions, knowledge of the rheological properties is required for the design, selection, and operation of the equipment involved. A vast amount of published literature exists on the rheology of dispersions of *rigid* particles [6–20]. However, dispersions of *soft* rubberlike solid particles have received less attention.

The rheological properties of dispersions of *soft* solid particles are more complicated than the rheological properties of *rigid*-particle dispersions because of the deformable nature of soft particles. For example, at very low concentrations of the dispersed phase, dispersions of spherical rigid particles exhibit Newtonian behavior, whereas dispersions of soft solid particles exhibit shear-thinning and viscoelastic properties [21–25]. Dispersions of soft solid particles also exhibit elastic recovery; they flow under the application of shear stress, but on removal of stress, part of their deformation is gradually recovered due to the elastic nature of the particles [21–25].

In this chapter, the macroscopic rheological behavior of dilute and concentrated dispersions of soft solid particles in liquids is reviewed.

II. RHEOLOGY OF DILUTE DISPERSIONS

For an infinitely dilute dispersion of *rigid* spheres in a Newtonian fluid, the average deviatoric stress tensor $\bar{\bar{\tau}}$ is linearly related to the average rate-of-strain tensor $\bar{\bar{d}}$ by the constitutive equation [11]:

$$\bar{\bar{\tau}} = 2\eta\bar{\bar{d}} \tag{1}$$

According to Eq. (1), an infinitely dilute dispersion of rigid spheres is a Newtonian fluid possessing a shear viscosity η.

Einstein [8,9] was the first to derive an equation for the viscosity η of an infinitely dilute suspension of *rigid* spheres uniformly dispersed in an incompressible Newtonian fluid of viscosity η_c. The Einstein equation is

$$\eta = \eta_c\left(1 + \frac{5}{2}\phi\right) \tag{2}$$

where ϕ is the volume fraction of particles. Einstein's analysis was based upon a calculation of the additional dissipation of mechanical energy caused by the introduction of a single spherical particle into a homogeneous shear flow.

The Newtonian constitutive equation [Eq. (1)] cannot be applied to dispersions of *soft* (*deformable*) solid particles. The first detailed calculation of the rheological properties of a dispersion of soft solid particles was made by Frohlich and Sack [21]. They utilized an effective medium approach to derive a constitutive equation governing the rheological behavior of a dilute dispersion of soft solid particles, at *small* variable rates of strain. According to the approach utilized by Frohlich and Sack, the dispersion is first treated as an equivalent "effective medium" which is homogeneous and has the same macroscopic rheological properties as that of the dispersion. Then, a small portion of the effective homogeneous medium is replaced by the actual components of the dispersion. The properties of the effective medium are then determined by insisting that if a small portion of the effective homogeneous medium is replaced by the actual components of the dispersion, no difference in rheological behavior could be detected by macroscopic observations. In essence, Frohlich and Sack [21] consider two systems: In one system, a single spherical particle of radius R, made up of elastic material with shear modulus G_p, is surrounded by an envelope of continuous-phase fluid (Newtonian of viscosity η_c) up to a certain radius a, the remaining space $r > a$ is filled with the effective homogeneous medium (dispersion) whose properties we need to calculate. The radius a is chosen to give the correct volume fraction of the dispersed phase [i.e., $a = R/(\phi)^{1/3}$]. In the other system, the whole space is filled with the effective homogeneous medium (dispersion) whose macro-

scopic properties we need to calculate. The two systems are then required to be indistinguishable when subjected to an arbitrary state of flow, provided that observations are made at a distance $r = b$ sufficiently large compared with a.

Frohlich and Sack's analysis leads to the following constitutive equation, valid at small variable rates of strain for dispersions of elastic spheres in Newtonian fluids:

$$\left(1 + \lambda_1 \frac{\partial}{\partial t}\right)\overline{\overline{\tau}} = 2\eta_0\left(1 + \lambda_2 \frac{\partial}{\partial t}\right)\overline{\overline{d}} \tag{3}$$

where λ_1 is a relaxation time, λ_2 is a retardation time, and η_0 is the zero-shear viscosity of dispersion. Expressions for η_0, λ_1, and λ_2 are found to be

$$\eta_0 = \eta_c\left(\frac{2 + 3\phi}{2 - 2\phi}\right) \tag{4}$$

$$\lambda_1 = \frac{\eta_c(3 + 2\phi)}{2G_p(1 - \phi)} \tag{5}$$

$$\lambda_2 = \frac{3\eta_c(1 - \phi)}{G_p(2 + 3\phi)} \tag{6}$$

Equation (3) is valid for dilute to moderately concentrated dispersions. For infinitely dilute dispersion, Eq. (3) is valid with the following expressions for η_0, λ_1, and λ_2:

$$\eta_0 = \eta_c\left(1 + \frac{5}{2}\phi\right) \tag{7}$$

$$\lambda_1 = \frac{3\eta_c}{2G_p}\left(1 + \frac{5}{3}\phi\right) \tag{8}$$

$$\lambda_2 = \frac{3\eta_c}{2G_p}\left(1 - \frac{5}{2}\phi\right) \tag{9}$$

Equations (7)–(9) follow from the more general equations [Eqs. (4)–(6)] when $\phi \to 0$.

In slow steady shearing, Frohlich and Sack's constitutive equation [Eq. (3)] reduces to the Newtonian constitutive equation [Eq. (1)]. However, in nonsteady experiments, the Frohlich and Sack constitutive equation predicts

viscoelasticity (delayed response to stress, elastic recovery upon removal of stress) in dispersions of soft solid particles.

The constitutive equation of Frohlich and Sack cannot be applied at finite rates of deformation. It is restricted to small (variable) rates of strain.

Goddard and Miller [25] derived a more general constitutive equation for an infinitely dilute dispersion of soft elastic particles in a Newtonian fluid. Their equation is

$$\overline{\overline{\tau}} + \lambda_0 \frac{D\overline{\overline{\tau}}}{D\mathrm{t}} = 2\eta_c \left[\left(1 + \frac{5}{2}\phi\right)\overline{\overline{d}} + \left(1 - \frac{5}{3}\phi\right)\lambda_0 \frac{D\overline{\overline{d}}}{Dt} + \frac{25}{7}\phi\lambda_0 \left\{ \overline{\overline{d}}\cdot\overline{\overline{d}} - \frac{1}{3}\,\mathrm{tr}\left(\overline{\overline{d}}\cdot\overline{\overline{d}}\right)\overline{\overline{\delta}} \right\} \right] \tag{10}$$

where D/Dt is the Jaumann derivative, tr refers to the trace of a tensor, $\overline{\overline{\delta}}$ is the unit tensor, and $\lambda_0 = 3\eta_c/2G_p$.

When a dispersion of soft elastic particles is subjected to steady shearing flow, Eq. (10) predicts the presence of both shear and normal stresses. It can be shown easily that for steady shearing flow, the Goddard and Miller constitutive equation gives the following expressions for the shear and normal stresses:

$$\tau_{12} = \frac{\eta_c\dot{\gamma}}{1 + \lambda_0^2\dot{\gamma}^2}\left[\left(1 + \frac{5}{2}\phi\right) + \lambda_0^2\dot{\gamma}^2\left(1 - \frac{5}{3}\phi\right)\right] \tag{11}$$

$$\tau_{11} - \tau_{22} = \frac{(25/3)\eta_c\lambda_0\dot{\gamma}^2\phi}{1 + \lambda_0^2\dot{\gamma}^2} \tag{12}$$

$$\tau_{22} - \tau_{33} = -\frac{\eta_c\lambda_0\dot{\gamma}^2\phi\left(50/21 - (25/24)\lambda_0^2\dot{\gamma}^2\right)}{1 + \lambda_0^2\dot{\gamma}^2} \tag{13}$$

where τ_{12} is the shear stress, $\dot{\gamma}$ is the shear rate, $\tau_{11} - \tau_{22}$ is first normal stress difference, and $\tau_{22} - \tau_{33}$ is second normal stress difference.

Thus, even an infinitely dilute dispersion of soft elastic particles exhibits shear-thinning behavior and nonzero normal stress differences. Note that Eq. (11) can be written in terms of viscosity η, defined as the ratio of shear stress (τ_{12}) to shear rate ($\dot{\gamma}$):

$$\eta = \frac{\tau_{12}}{\dot{\gamma}} = \frac{\eta_c}{1 + \lambda_0^2\dot{\gamma}^2}\left[\left(1 + \frac{5}{2}\phi\right) + \lambda_0^2\dot{\gamma}^2\left(1 - \frac{5}{3}\phi\right)\right] \tag{14}$$

At low shear rates ($\dot{\gamma} \to 0$), Eq. (14) reduces to the Einstein formula [Eq. (2)] for the viscosity of dilute dispersion of *rigid* spheres.

A. Linear Viscoelastic Behavior of Dilute Dispersions

The linear viscoelastic behavior of materials is often described in terms of the response of a material to small-amplitude oscillatory shear. The important functions in oscillatory shear characterization are storage modulus (G') and loss modulus (G''). The storage modulus (G') gives information about the elastic character of the material or the energy storage that takes place during the deformation. The loss modulus (G''), on the other hand, gives information about the viscous character of the material or the energy dissipation that occurs during the deformation. The complex shear modulus (G^*) of the material is defined as

$$G^* = G' + jG'' \tag{15}$$

where j is the imaginary number $\left(\sqrt{-1}\right)$, and the complex shear viscosity (η^*) of the material is defined as

$$\eta^* = \eta' - j\eta'' \tag{16}$$

where η' is called dynamic viscosity and η'' is the imaginary part of complex viscosity. Note that

$$\eta^* = \frac{G^*}{j\omega} \tag{17}$$

$$\eta' = \frac{G''}{\omega} \tag{18}$$

$$\eta'' = \frac{G'}{\omega} \tag{19}$$

where ω is the frequency of oscillation of stress or strain.

For an infinitely dilute dispersion of rigid spheres in a Newtonian fluid, the storage modulus (G') and the imaginary part of complex viscosity (η'') are zero, whereas the loss modulus (G'') and dynamic viscosity (η') are given as

$$G'' = \eta'\omega = \eta_c\left(1 + \frac{5}{2}\phi\right)\omega \tag{20}$$

This result follows from the Einstein formula [Eq. (2)].

Unlike a dilute dispersion of *rigid* spheres, a dilute dispersion of *soft* (elastic) solid spheres exhibits nonzero values of storage modulus and η'' because of its ability to store energy during the deformation.

For dilute to moderately concentrated dispersions of soft (elastic) solid spheres, an expression for the complex shear viscosity can be derived from the

Frohlich and Sack constitutive equation [Eq. (3)]. If, in Eqs. (3)–(6), η is replaced by η^* (complex viscosity), η_c by η_c^*, G_p by G_d^* (complex modulus of dispersed phase), and $\partial/\partial t$ by $j\omega$, the following equation is obtained:

$$\frac{\eta^*}{\eta_c^*} = \left(\frac{2G_d^* + 3j\omega\eta_c^* + 3\phi(G_d^* - j\omega\eta_c^*)}{2G_d^* + 3j\omega\eta_c^* - 2\phi(G_d^* - j\omega\eta_c^*)}\right) \tag{21}$$

Because $G^* = j\omega\eta^*$, we can rewrite Eq. (21) in terms of complex shear modulus as

$$\frac{G^*}{G_c^*} = \left(\frac{2G_d^* + 3G_c^* + 3\phi(G_d^* - G_c^*)}{2G_d^* + 3G_c^* - 2\phi(G_d^* - G_c^*)}\right) \tag{22}$$

This equation for complex shear modulus is valid for dilute to moderately concentrated dispersions of soft viscoelastic particles (modulus G_d^*) in a viscoelastic matrix (modulus G_c^*). In the limit $\phi \to 0$ (i.e., infinitely dilute systems), Eq. (22) reduces to

$$\frac{G^*}{G_c^*} = 1 + \frac{5}{2}\phi\left(\frac{2(G_d^* - G_c^*)}{2G_d^* + 3G_c^*}\right) \tag{23}$$

For dispersions of rigid spherical particles ($G_d^* \to \infty$) in a Newtonian liquid ($G_c^* = j\omega\eta_c$), Eq. (23) gives the Einstein formula [Eq. (2)] for the viscosity of infinitely dilute dispersion of rigid spheres in a Newtonian fluid.

III. RHEOLOGY OF CONCENTRATED DISPERSIONS

The constitutive equation of Goddard and Miller [Eq. (10)] cannot be applied at finite concentrations of dispersed phase (elastic particles) because the hydrodynamic interaction between the particles is ignored in its derivation. Equation (10) is probably valid for ϕ less than about 0.02.

The Frohlich and Sack constitutive equation [Eq. (3)] is valid for moderately concentrated dispersions ($\phi \leq 0.1$) of soft elastic particles. However, one serious limitation of this equation is that it is restricted to low rates of strain. Oldroyd [23] has suggested that the Frohlich and Sack constitutive equation can be applied to finite rates of strain provided that $\partial/\partial t$ in Eq. (3) is replaced by D/Dt, the Jaumann derivative. Thus, the modified Frohlich and Sack constitutive equation is

$$\left(1 + \lambda_1 \frac{D}{Dt}\right)\overline{\overline{\tau}} = 2\eta_0\left(1 + \lambda_2 \frac{D}{Dt}\right)\overline{\overline{d}} \tag{24}$$

where η_0, λ_1, and λ_2 are given by Eqs. (4)–(6).

In steady simple shearing motion, Eq. (24) gives the following expressions for the viscosity and normal stress differences:

$$\eta = \frac{\tau_{12}}{\dot{\gamma}} = \frac{\eta_0\left(1 + \lambda_1\lambda_2\dot{\gamma}^2\right)}{1 + \lambda_1^2\dot{\gamma}^2} \tag{25}$$

$$\tau_{11} - \tau_{22} = \frac{2\eta_0(\lambda_1 - \lambda_2)\dot{\gamma}^2}{1 + \lambda_1^2\dot{\gamma}^2} \tag{26}$$

$$\tau_{22} - \tau_{33} = -\frac{\eta_0(\lambda_1 - \lambda_2)\dot{\gamma}^2}{1 + \lambda_1^2\dot{\gamma}^2} \tag{27}$$

where η_0, λ_1, and λ_2 are given by Eqs. (4)–(6). Like the Goddard and Miller constitutive equation [Eq. (10)], the modified Frohlich and Sack constitutive equation [Eq. (24)] also predicts shear-thinning behavior and nonzero normal stress differences for dispersions of soft elastic particles in a Newtonian fluid.

A. New Viscosity Models for Concentrated Dispersions

The viscosity equation of Frohlich and Sack [Eq. (25)] is applicable only for moderately concentrated dispersions ($\phi \leq 0.1$) of soft elastic particles. It generally underpredicts the viscosity at high concentrations of the dispersed phase ($\phi > 0.1$). For instance, the Frohlich and Sack viscosity equation predicts that the viscosity of dispersions of *rigid* spheres ($G_p \to \infty$) is

$$\eta = \eta_c\left(\frac{2 + 3\phi}{2(1 - \phi)}\right) = \eta_c\left(1 + \frac{5}{2}\phi + \frac{5}{2}\phi^2 + \cdots\right) \tag{28}$$

This equation gives the coefficient of ϕ^2 to be 2.5, whereas the actual value determined by Batchelor and Green [19] is about two times this value. Furthermore, the Frohlich and Sack viscosity equation fails to exhibit divergence in dispersion viscosity η at $\phi \to \phi_m$, where ϕ_m is the maximum packing volume fraction of particles. At $\phi \to \phi_m$, η is expected to diverge especially for dispersions of rigid particles (with $G_p \to \infty$). The Frohlich and Sack viscosity equation predicts divergence in η only at $\phi = 1.0$ [see Eq. (28)].

Pal [26] has recently derived new equations for the viscosity of concentrated suspensions of deformable elastic particles using the differential effective medium approach (DEMA). According to this approach, a concentrated dispersion is considered to be obtained from an initial continuous phase by successively adding infinitesimally small quantities of particles to the system while the final volume fraction of the dispersed phase is reached. At any arbitrary stage (i) of the process, the addition of an infinitesimal amount

of particles leads to the next stage $i + 1$. The dispersion of stage i is then treated as an equivalent "effective medium," which is homogeneous with respect to the new set of particles added to reach stage $i + 1$. The solution of a dilute dispersion is then applied to determine the increment increase in viscosity in going from stage i to stage $i + 1$. The differential equation derived in this manner is integrated to obtain the final solution for a concentrated dispersion. This DEMA was originally proposed by Bruggeman [27], who derived an equation for the dielectric constant of a concentrated dispersion. The DEMA is often referred to as "Brinkman–Roscoe" differential scheme in the rheology literature because Brinkman [18] and Roscoe [17] were the first to apply this scheme for the derivation of a viscosity equation for concentrated dispersion of *rigid* spherical particles. Brinkman [18] and Roscoe [17] utilized the celebrated Einstein viscosity equation [Eq. (2)] as a solution of dilute dispersion in the derivation of their viscosity equation for concentrated dispersions.

Pal [26] applied the DEMA to derive the viscosity equations for concentrated suspensions of deformable elastic particles. The Goddard and Miller viscosity equation [Eq. (14)] was used as a solution of dilute dispersion of deformable elastic particles. The Goddard and Miller viscosity equation [Eq. (14)] can be written in the form

$$\eta_r = 1 + \phi\left(\frac{[\eta]\left(1 - \frac{3}{2}N_{se}^2\right)}{1 + \left(\frac{3}{2}N_{se}\right)^2}\right) \tag{29}$$

where η_r is the relative viscosity, defined as the ratio of dispersion viscosity (η) to continuous medium viscosity (η_c), ϕ is the volume fraction of the dispersed phase, $[\eta]$ is the intrinsic viscosity (equal to 2.5 for rigid spherical particles), N_{se} is the shearoelastic number, given by

$$N_{se} = \frac{\eta_c \dot{\gamma}}{G_p} \tag{30}$$

Pal's models for the viscosity of concentrated dispersions of deformable elastic particles are summarized as follows:

Model 1:

$$\eta_r\left(\frac{1 - \frac{3}{2}\eta_r^2 N_{se}^2}{1 - \frac{3}{2}N_{se}^2}\right)^{-5/4} = \exp([\eta]\phi) \tag{31}$$

Model 2:

$$\eta_r\left(\frac{1 - \frac{3}{2}\eta_r^2 N_{se}^2}{1 - \frac{3}{2}N_{se}^2}\right)^{-5/4} = (1 - \phi)^{-[\eta]} \tag{32}$$

Model 3:

$$\eta_r \left(\frac{1 - \frac{3}{2}\eta_r^2 N_{se}^2}{1 - \frac{3}{2} N_{se}^2} \right)^{-5/4} = \exp\left[\frac{[\eta]\phi}{1 - \phi/\phi_m} \right] \tag{33}$$

Model 4:

$$\eta_r \left(\frac{1 - \frac{3}{2}\eta_r^2 N_{se}^2}{1 - \frac{3}{2} N_{se}^2} \right)^{-5/4} = \left(1 - \frac{\phi}{\phi_m} \right)^{-[\eta]\phi_m} \tag{34}$$

Pal's models [models 1–4, Eqs. (31)–(34)] predict shear-thinning behavior, as expected for dispersions of deformable elastic particles; the relative viscosity η_r decreases with the increase in shearoelastic number N_{se}. Models 3 [Eq. (33)] and 4 [Eq. (34)] exhibit divergence of viscosity at $\phi = \phi_m$. Furthermore, models 3 and 4 are capable of taking into account the effect of particle size distribution on the viscosity of dispersions, through the parameter ϕ_m. Note that ϕ_m is sensitive to particle size distribution; for example, when a monodisperse system is changed to a bimodal dispersion, an increase in ϕ_m is expected.

Pal's models can be written in a more general form as

$$\eta_r \left(\frac{1 - \frac{3}{2}\eta_r^2 N_{se}^2}{1 - \frac{3}{2} N_{se}^2} \right)^{-5/4} = f([\eta], \phi, \phi_m) \tag{35}$$

In the limit $N_{se} \to 0$, the left-hand term of Eq. (35) becomes η_r. Consequently, Eq. (35) reduces to

$$\eta_r = f([\eta], \phi, \phi_m) \tag{36}$$

An important point to note is that the function f ([η], ϕ, ϕ_m) is the same for dispersions of *soft* elastic particles and for dispersions of *rigid* particles. This implies that we can transform relative viscosity–concentration equations for dispersions of *rigid* particles to corresponding equations for dispersions of *soft* elastic particles by replacing η_r with $\eta_r\left(1 - 3/2\ \eta_r^2 N_{se}^2\right)^{-5/4}\left(1 - 3/2\ N_{se}^2\right)^{5/4}$.

B. New Linear Viscoelastic Models for Concentrated Dispersions

The complex shear modulus equation of Frohlich and Sack [Eq. (22)] is valid for dilute to moderately concentrated dispersions of deformable particles. When the dispersed particles are elastic $\left(G_d^* = G_p\right)$ and the continuous phase (matrix) is Newtonian $\left(G_c^* = j\omega\eta_c\right)$, Eq. (22) can be written as

$$G^* = j\omega\eta_0 \left(\frac{1 + j\omega\lambda_2}{1 + j\omega\lambda_1} \right) \tag{37}$$

where η_0, λ_1, and λ_2 are given by Eqs. (4)–(6). Equation (37) gives the following expressions for storage and loss moduli:

$$G' = \frac{\eta_0 \omega^2 (\lambda_1 - \lambda_2)}{1 + \omega^2 \lambda_1^2} \tag{38}$$

$$G'' = \frac{\eta_0 \omega \left(1 + \omega^2 \lambda_1 \lambda_2\right)}{1 + \omega^2 \lambda_1^2} \tag{39}$$

One serious limitation of the Frohlich and Sack complex modulus equation [Eq. (22)] is that it fails to predict accurate values of the storage and loss moduli at high concentrations of dispersed phase ($\phi > 0.20$); Equation (22) generally underpredicts G' and G'' at high ϕ. Furthermore, it fails to exhibit a divergence in G^* at $\phi = \phi_m$. At $\phi = \phi_m$, G^* is expected to diverge, especially for dispersions with $G_d^* \to \infty$. Note that in the limit $G_d^* \to \infty$, Eq. (22) reduces to

$$G^* = G_c^* \left(\frac{2 + 3\phi}{2(1 - \phi)} \right) \tag{40}$$

Equation (40) predicts divergence in G^* only at $\phi = 1.0$. The maximum packing volume fraction of particles ϕ_m where divergence in G^* is expected is usually much smaller than 1.0.

Pal [28] has recently derived new equations for the complex shear modulus of concentrated dispersions of deformable elastic particles using the DEMA, discussed earlier. Using Eq. (23) as a solution for the complex shear modulus of dilute dispersion of soft deformable particles, the following equations for concentrated dispersions were derived by applying DEMA:

Model 1:

$$\left(\frac{G^*}{G_c^*} \right) \left(\frac{G^* - G_d^*}{G_c^* - G_d^*} \right)^{-2.5} = \exp(2.5\phi) \tag{41}$$

Model 2:

$$\left(\frac{G^*}{G_c^*} \right) \left(\frac{G^* - G_d^*}{G_c^* - G_d^*} \right)^{-2.5} = (1 - \phi)^{-2.5} \tag{42}$$

Model 3:

$$\left(\frac{G^*}{G_c^*} \right) \left(\frac{G^* - G_d^*}{G_c^* - G_d^*} \right)^{-2.5} = \exp\left(\frac{2.5\phi}{1 - \phi/\phi_m} \right) \tag{43}$$

Model 4:

$$\left(\frac{G^*}{G_c^*}\right)\left(\frac{G^* - G_d^*}{G_c^* - G_d^*}\right)^{-2.5} = \left(1 - \frac{\phi}{\phi_m}\right)^{-2.5\phi_m} \tag{44}$$

Pal's models for the complex shear modulus of concentrated dispersions of soft viscoelastic particles can be written in a more general form as

$$\left(\frac{G^*}{G_c^*}\right)\left(\frac{G^* - G_d^*}{G_c^* - G_d^*}\right)^{-2.5} = f(\phi, \phi_m) \tag{45}$$

In the limit $G_d^* \to \infty$, Eq. (45) reduces to

$$\frac{G^*}{G_c^*} = f(\phi, \phi_m) \tag{46}$$

Because the function $f(\phi, \phi_m)$ is the same for dispersions of soft and rigid particles, one can transform complex modulus equations for dispersions of *rigid* particles to corresponding equations for dispersions of *soft* particles by replacing G^*/G_c^* by $(G^*/G_c^*)\,(G^* - G_d^*)^{-2.5}\,(G_c^* - G_d^*)^{2.5}$.

IV. MODEL PREDICTIONS AND COMPARISONS WITH EXPERIMENTAL DATA

A. Model Predictions

Figure 1 shows the effects of particle rigidity and matrix (continuous-phase) viscosity on the relative viscosity of a concentrated dispersion ($\phi = 0.45$) of soft deformable particles. The plots are generated from Pal's model 4 [Eq. (34)] with $[\eta] = 2.5$ and $\phi_m = 0.637$. In the top plot of Fig. 1, relative viscosity is shown as a function of shear stress ($= \eta_c\dot{\gamma}$) acting on the particles, for different values of particle shear modulus G_p. The rigid particle dispersion ($G_p \to \infty$) behaves as a Newtonian fluid. With the decrease in G_p, shear-thinning behavior is observed. The bottom plot of Fig. 1 shows relative viscosity as a function of shear rate $\dot{\gamma}$ for different values of the continuous-medium viscosity η_c. With the increase in η_c, the relative viscosity–shear rate curve shifts to the left. This indicates that the shear-thinning behavior in dispersion begins at a lower shear rate when η_c is increased. In the shear-rate range where shear thinning is present in the dispersion, the relative viscosity decreases with the increase in η_c. The enhancement of shear-thinning behavior and decrease in the relative viscosity exhibited by the dispersion, when η_c is increased and G_p is decreased, can be explained in terms of an increase in the deformation of particles.

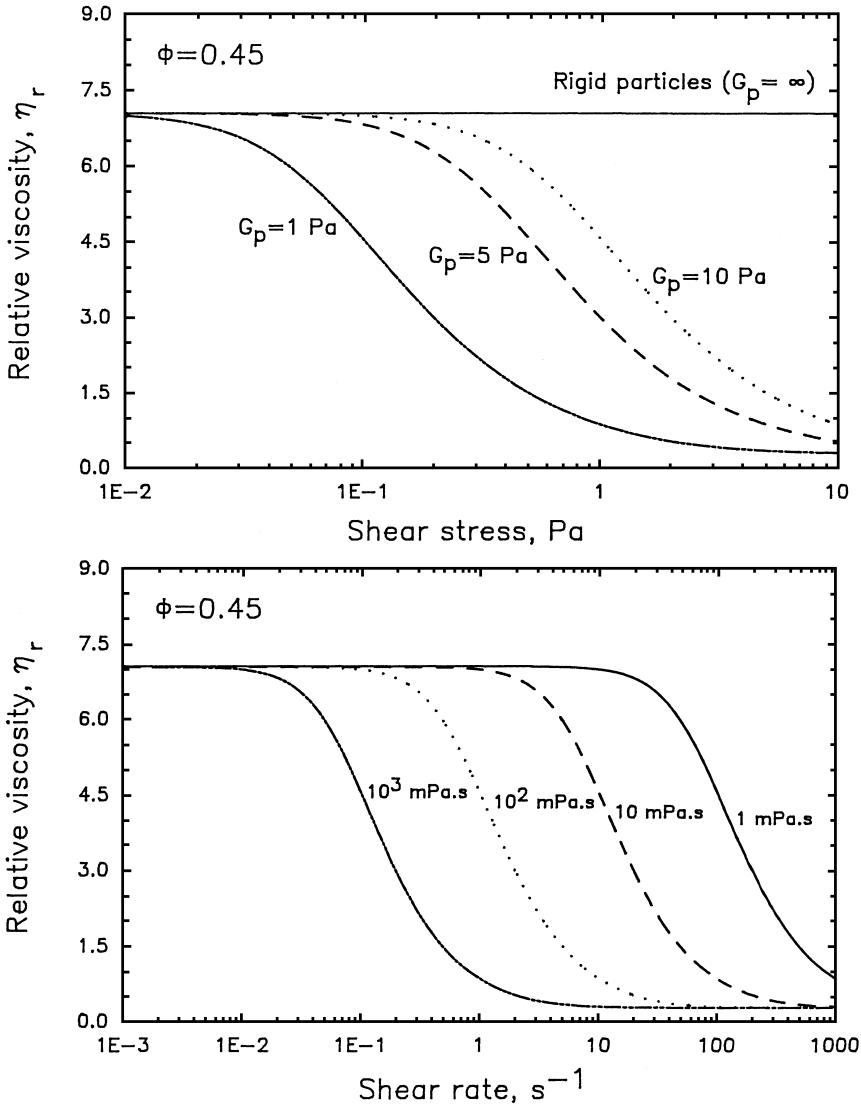


Figure 1 Effects of particle rigidity and continuous-medium viscosity on the relative viscosity of a concentrated dispersion. The plots are generated from Pal's model [Eq. (34)] using $[\eta] = 2.5$ and $\phi_m = 0.637$. G_p is the particle shear modulus.

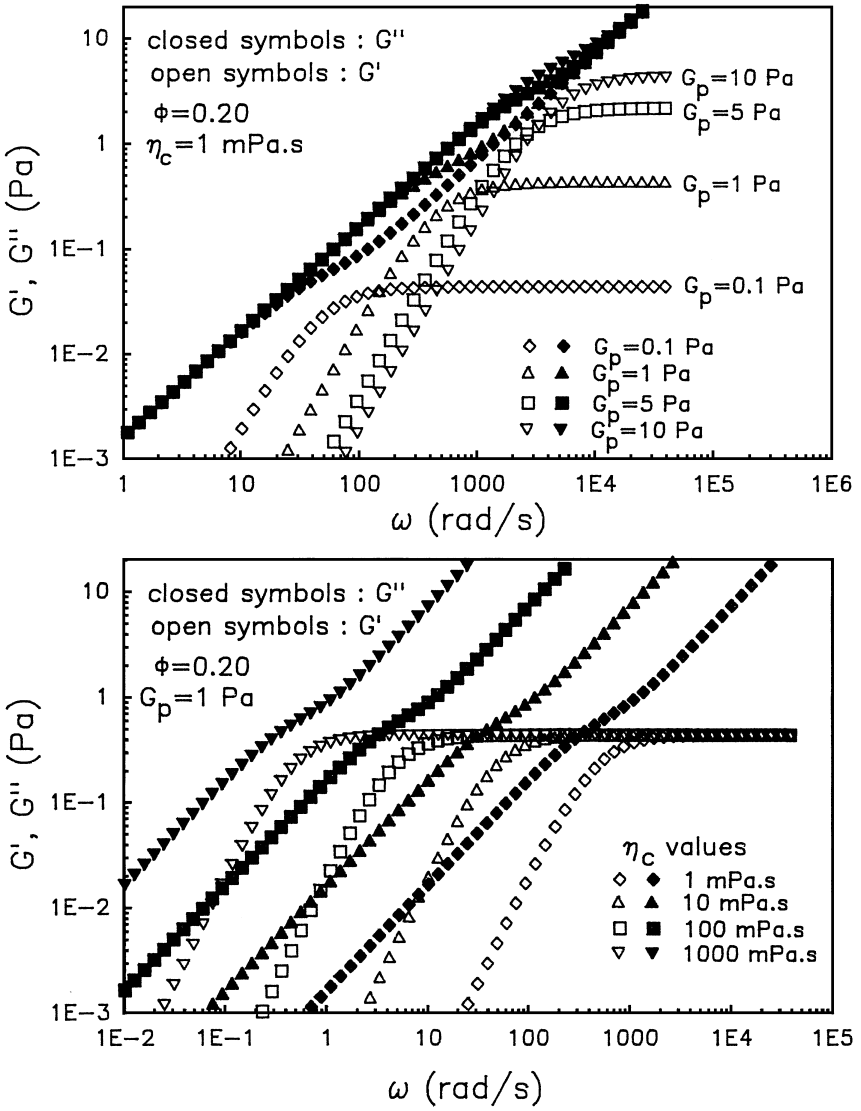


Figure 2 Effects of particle rigidity and continuous-medium viscosity on the storage and loss moduli of a dispersion. The plots are generated from the Frohlich and Sack model [Eqs. (38) and (39)].

Figure 2 shows the effects of particle rigidity and matrix (continuous-phase) viscosity on the storage and loss moduli of a dispersion of soft elastic particles. The plots are generated from the Frohlich and Sack model [Eqs. (38) and (39)]. In the top plot of Fig. 2, storage modulus (G') and loss modulus (G'') versus frequency are shown, for different values of particle shear modulus G_p. With the increase in particle rigidity G_p, the plateau value of the storage modulus (G') of the dispersion increases. Also, the frequency at which the terminal zone ($G' \alpha \omega^2$) begins is shifted to higher values with increase in G_p. This is because the relaxation time of particles decreases with an increase in G_p. Particle rigidity appears to have only a small effect on the loss modulus (G'') of the dispersion. The bottom plot of Fig. 2 shows G' and G'' versus frequency ω for different values of the continuous-medium viscosity η_c. With the increase in η_c, the plateau value of the storage modulus of the dispersion is unaffected. However, the frequency at which the terminal zone begins is shifted to lower values with an increase in η_c. This is because the relaxation time of particles increases with an increase in η_c. The loss modulus–frequency plots shift toward the left with the increase in η_c, indicating an increase in G''.

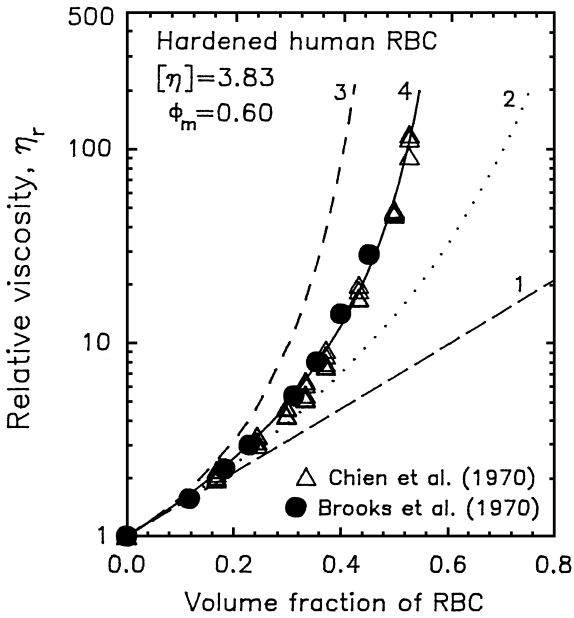


Figure 3 Relative viscosity (η_r) as a function of volume fraction of hardened RBCs. Comparison between experimental data and predictions of the Pal models. Numbers shown on different curves refer to model number.

B. Comparison with Experimental Data

Pal's models [Eqs. (31)–(34)] for the viscosity of concentrated dispersions of soft elastic particles are evaluated using experimental data on dispersions of normal and hardened human erythrocytes [red blood cells (RBCs)]. According to Pal [26], it is quite appropriate to treat RBCs as deformable elastic particles. However, it should be noted that RBCs are not spherical particles; for nonspherical particles such as RBCs, the intrinsic viscosity $[\eta]$ is expected to be greater than 2.5.

Figure 3 shows the comparison between the experimental data of hardened RBC dispersions and predictions of the Pal models. As can be seen, the experimental data of both Chien et al. [29] and Brooks et al. [30] for dispersions of hardened RBCs can be described very well with model 4 of Pal [Eq. (34)] using $[\eta] = 3.83$ and $\phi_m = 0.60$. Model 3 [Eq. (33)] overpredicts the relative viscosity, whereas models 2 [Eq. (32)] and 1 [Eq. (31)] underpredict relative viscosity. Note that for hardened RBCs, $G_p \rightarrow \infty$ and $N_{se} \rightarrow 0$ and, therefore, dispersions of hardened RBCs exhibit Newtonian behavior (viscosity independent of shear rate).

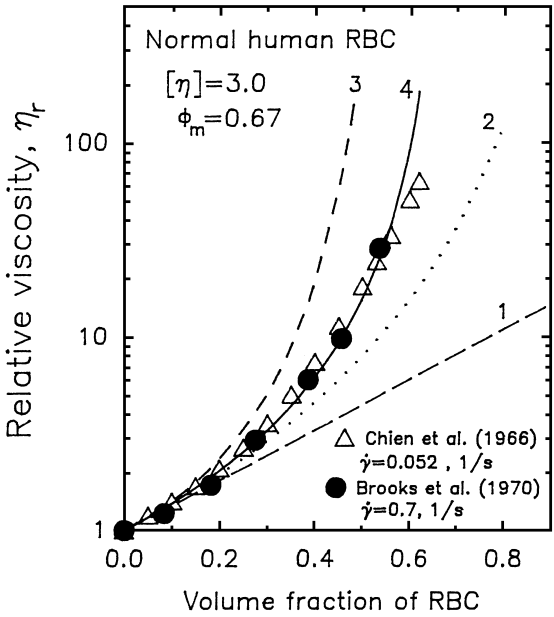


Figure 4 Relative viscosity (η_r) as a function of volume fraction of normal RBCs at low shear rates. Comparison between experimental data and predictions of various Pal models. Numbers shown on different curves refer to model number.

Figure 4 shows the comparison between the experimental data of normal RBC dispersions at low shear rate ($N_{se} \rightarrow 0$) and predictions of the Pal models. The experimental data of both Chien et al. [31] and Brooks et al. [30] for dispersions of normal RBCs at low shear rate can be described very well with model 4 of Pal [Eq. (34)] using [η] and ϕ_m values of 3.0 and 0.67, respectively. As in the case of hardened RBC dispersions, model 3 [Eq. (33)] overpredicts η_r, whereas models 2 [Eq. (32)] and 1 [Eq. (31)] underpredict η_r.

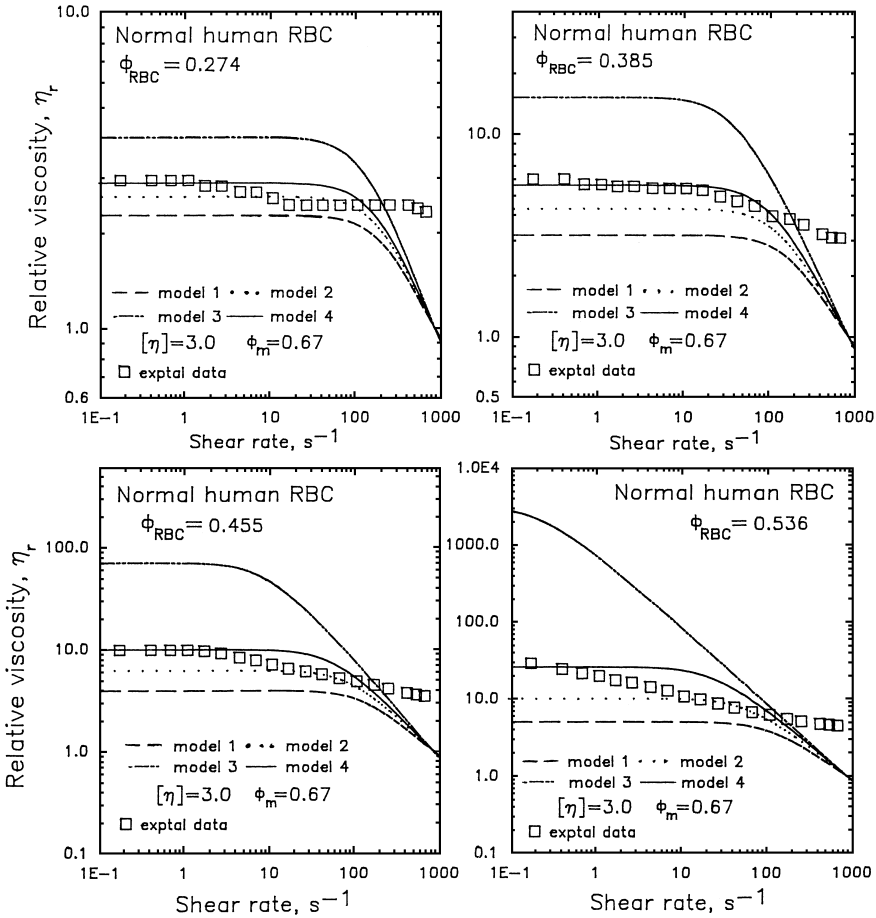


Figure 5 Relative viscosity (η_r) as a function of shear rate. Comparison between experimental data of Brooks et al. [30] on normal RBC dispersions and predictions of various Pal models.

Figure 5 shows the experimental η_r data of Brooks et al. [30] as a function of shear rate for normal RBC dispersions. Dispersions of normal RBCs exhibit shear-thining behavior; this shear-thinning behavior is expected for dispersions of soft elastic particles, such as normal RBCs. The predictions of the Pal models are also shown in Fig. 5. The shear modulus of particles (RBCs) was taken to be 1 Pa (i.e., $G_p = 1$ Pa). This value of G_p corresponds to the ratio of elastic shear modulus of the RBC membrane to the average radius of the RBC [26]. The experimental data of Brooks et al. show reasonably good agreement with model 4 [Eq. (34)] predictions when shear rates are less than 200 s^{-1}.

Pal's models [Eqs. (41)–(44)] for the complex modulus of concentrated dispersions of elastic particles are evaluated using experimental data on dispersions of spherical glass beads ($G_p \to \infty$) in a polymeric liquid [28]. Figures 6–9 compare experimental data with the predictions of the Pal models. The experimental data, consisting of phase lag angle and complex

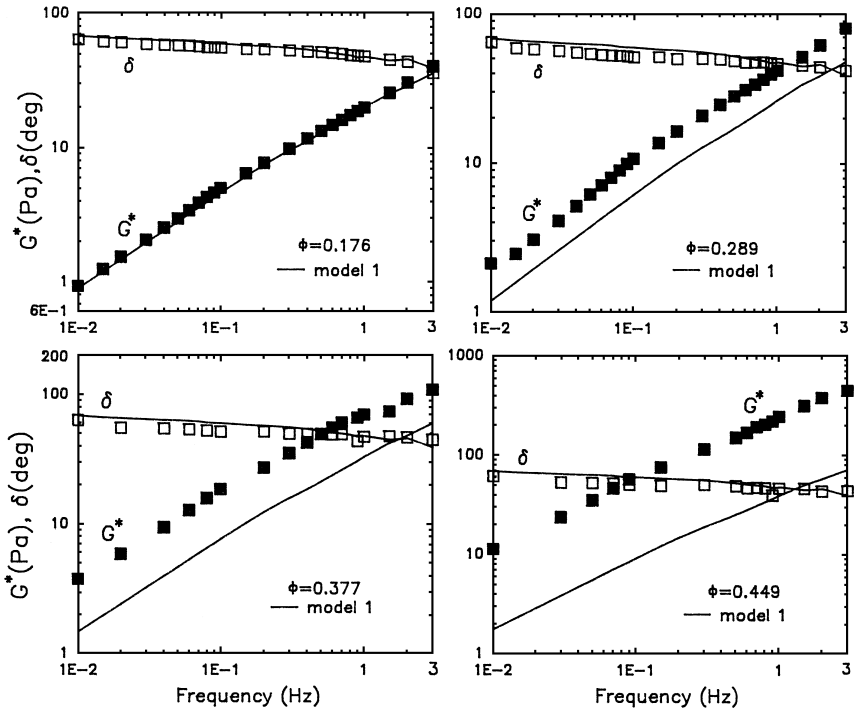


Figure 6 Comparison of complex shear modulus/phase lag angle data of glass bead dispersions with predictions of Pal's model 1 [Eq. (41)].

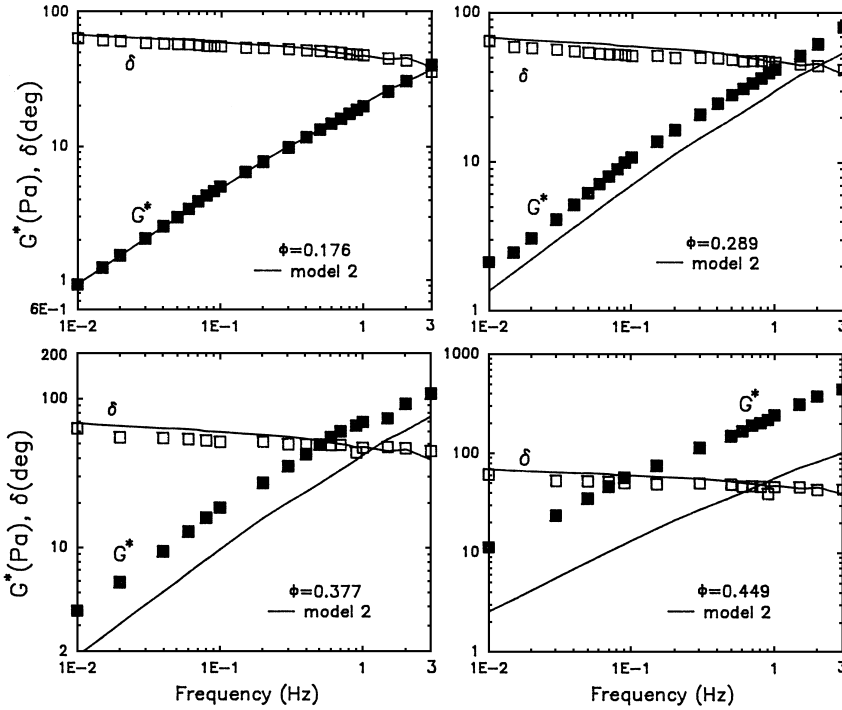


Figure 7 Comparison of complex shear modulus/phase lag angle data of glass bead dispersions with predictions of Pal's model 2 [Eq. (42)].

modulus as functions of oscillation frequency of stress, are plotted and compared with model predictions at various volume fractions of the dispersed phase.

Model 1 [Eq. (41)] gives good predictions of the complex modulus only at low volume fractions of the dispersed phase ($\phi \leq 0.176$). At higher values of ϕ, model 1 underpredicts the complex modulus and the deviation increases with the increase in ϕ value.

Model 2 [Eq. (42)] is a slight improvement over model 1. At low volume fractions of the dispersed phase ($\phi \leq 0.176$), model 2 gives good predictions of the complex modulus. At higher values of ϕ, model 2, like model 1, underpredicts the complex modulus.

Model 3 [Eq. (43)] gives excellent predictions of the complex modulus over the full range of ϕ, covered in the experimental work ($0 \leq \phi \leq 0.449$) provided that the ϕ_m value of 0.74 is used.

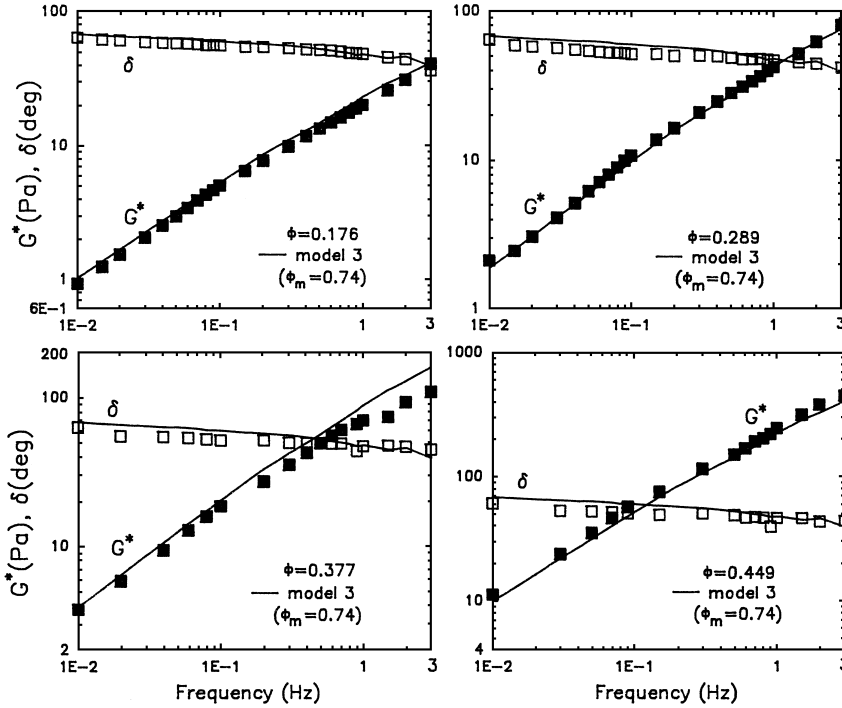


Figure 8 Comparison of complex shear modulus/phase lag angle data of glass bead dispersions with predictions of Pal's model 3 [Eq. (43)].

Model 4 [Eq. (44)] also gives excellent predictions of the complex modulus over the full range of ϕ covered when the ϕ_m value of 0.50 is used. Whereas model 3 gives a high ϕ_m value of 0.74, corresponding to hexagonal close-packed arrangement of uniform spheres, model 4 gives a significantly lower value of ϕ_m, corresponding to simple cubic packing of uniform spheres.

Interestingly, all four models of Pal give good prediction of the phase lag angle (δ). As expected from the models, δ for dispersions is nearly the same as that of the matrix (continuous phase).

V. CONCLUDING REMARKS

Dispersions of soft rubberlike (elastic) particles exhibit complicated rheological properties such as shear thinning in steady shear, nonzero normal stress differences in steady shear, and nonzero storage modulus in oscillatory shear.

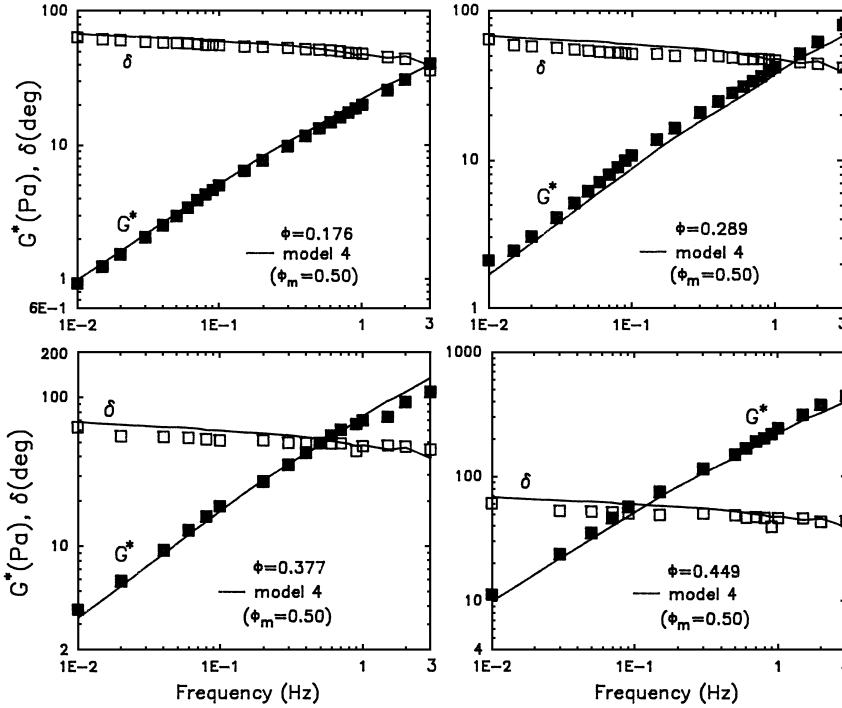


Figure 9 Comparison of complex shear modulus/phase lag angle data of glass bead dispersions with predictions of Pal's model 4 [Eq. (44)].

The non-Newtonian rheological behavior exhibited by dispersions of soft particles is due to the deformable nature of particles; dispersions of rigid particles exhibit Newtonian behavior, especially at low to moderate concentrations of dispersed particles. The rheological models for dilute and concentrated dispersions of soft particles are reviewed. The model predictions are compared with experimental data available on dispersions. Only a limited amount of experimental data is available in the literature. Further experimental work is needed on the rheology of dispersions of soft solid particles.

ACKNOWLEDGMENTS

Financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada is appreciated.

REFERENCES

1. Tadros, Th.F. Ed. *Solid/liquid Dispersions*; Academic Press: London, 1987.
2. Kissa, E. *Dispersions: Characterization, Testing, and Measurement*; Marcel Dekker: New York, 1999.
3. Han, C.D. *Multiphase Flow in Polymer Processing*; Academic Press: New York, 1981.
4. Conley, R.F. *Practical Dispersion*; VCH: New York, 1996.
5. Nielloud, F.; Marti-Mestres, G. Eds. *Pharmaceutical Emulsions and Suspensions*; Marcel Dekker: New York, 2000.
6. Gupta, R.K. *Polymer and Composite Rheology*; Marcel Dekker: New York, 2000.
7. Larson, R.G. *The Structure and Rheology of Complex Fluids*; Oxford University Press: New York, 1999.
8. Einstein, A. Ann. Phys. 1906, *19*, 289–306.
9. Einstein, A. Ann. Phys. 1911, *34*, 591–592.
10. Jeffery, D.J.; Acrivos, A. AIChE J. 1976, *22*, 417–432.
11. Brenner, H. *In Progress in Heat and Mass Transfer*; Schowalter, W.R. Minkowycz, W.J., Luikov, A.V., Afgan, N.H. Eds.; Pergaman Press: Oxford, 1972; Vol. 5, 89–129.
12. Mill, C.C., Ed. *Rheology of Disperse Systems*; Pergaman Press: London, 1959.
13. Mooney, M. J. Colloid Sci. 1951, *6*, 162–170.
14. Krieger, I.M.; Dougherty, T.J. Trans. Soc. Rheol. 1959, *3*, 137–152.
15. Krieger, I.M. Adv. Colloid. Interf. Sci. 1972, *3*, 111–136.
16. Chong, J.S.; Christiansen, E.B.; Baer, A.D. J. Appl. Polym. Sci. 1971, *15*, 2007–2021.
17. Roscoe, R. Br. J. Appl. Phys. 1952, *3*, 267–269.
18. Brinkman, H.C. J. Chem. Phys. 1952, *20*, 571.
19. Batchelor, G.K.; Green, J.T. J. Fluid Mech. 1972, *56*, 401–427.
20. Macosko, C.W. *Rheology Principles, Measurements, and Applications*; VCH: New York, 1994; 425–474.
21. Frohlich, H.; Sack, R. Proc. Roy. Soc. London A 1946, *185*, 415–430.
22. Oldroyd, J.G. Proc. Roy. Soc. London A 1953, *218*, 122–132.
23. Oldroyd, J.G. *In Rheology of Disperse Systems*; Mill, C.C., Ed.; Pergamon Press: London, 1959; 1–150.
24. Roscoe, R. J. Fluid Mech. 1967, *28*, 273–293.
25. Goddard, J.D.; Miller, C. J. Fluid Mech. 1967, *28*, 657–673.
26. Pal, R. J. Biomechanics. 2003, *36*, 981–989.
27. Bruggeman, D.A.G. Ann Phys 1935, *24*, 636–679.
28. Pal, R. J. Colloid Interf. Sci. 2002, *245*, 171–177.
29. Chien, S.; Usami, S.; Dellenback, R.J.; Gregersen, M.I. Am. J. Physiol. 1970, *219*, 136–142.
30. Brooks, E.E.; Goodwin, J.W.; Seaman, G.V.F. J. Appl. Physiol. 1970, *28*, 172–177.
31. Chien, S.; Usami, S.; Taylor, H.M.; Lundberg, J.L.; Gregersen, M.I. J. Appl. Physiol. 1966, *21*, 81–87.

8

Computer Simulations of Mechanical Micromanipulation of Proteins

ALEXANDRE S. LEMAK and JAMES R. LEPOCK University of Toronto, Toronto, Ontario, Canada

JEFF Z. Y. CHEN University of Waterloo, Waterloo, Ontario, Canada

I. INTRODUCTION

Molecular biophysics has become an interdisciplinary area that deals with a number of important fundamental issues in biology, chemistry, and physics. Although researchers from different disciplines often focus on the physical aspects of biopolymers that are relevant to their own field, the integration of experimental and computational techniques from different research areas has provided new tools that were unavailable a decade ago to study biopolymers at a microscopic scale.

One example of such success is the study of the conformation of a stretched DNA molecule in an elongation flow [1,2]. Schematically outlined in Fig. 1a, in an elongational flow the fluid displays a flow pattern that produces a velocity gradient near the central region. When a biopolymer is placed in this region, it experiences a stretching force in the horizontal direction and a compressing force in the vertical direction. Smith and co-workers have demonstrated an experiment of observing how a long DNA molecule is stretched in an elongational flow by cleverly labeling the molecule with fluorescence, making direct observation of the conformation possible [1,2]. They have stretched a double-strand DNA (dsDNA) to about a micrometer with an elongational rate of about 0.86 s^{-1}.

The second example of mechanically stretching a biopolymer is the use of an atomic force microscope (AFM), which is now commercially available and becoming a common tool, to directly pull the end of a biopolymer [3–6]. The basic idea is to examine how the internally bonded structure is ripped apart, to relate the breakage of the bonds to the required external force on

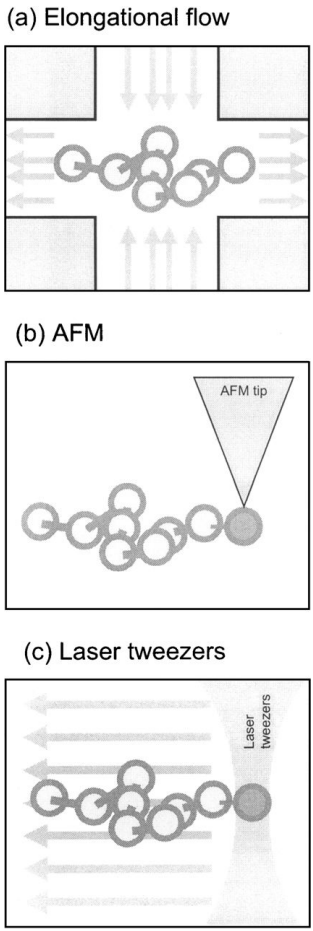


Figure 1 Schematic view of mechanical manipulation experiments: (a) a molecule in an elongational flow; (b) a molecule being pulled by the atomic force microscope tip that picks one monomer, at the terminal end as illustrated, or at some other points in the middle (not illustrated); (c) a molecule being stretched by a uniform flow while one terminal end is fixed in space by an optical trapping device.

the AFM tip (Fig. 1b). The AFM tip is usually charged so that it can effectively bond to the terminal end (or even the middle) of a biopolymer. The other end of the molecule can be fixed in space or left free depending on the system. Carrion-Vazques and co-workers, for example, have stretched a relatively large protein, the human cardiac titin, to a maximum extension of 200 nm with a force in the 200 pN range.

Yet another method of stretching a biopolymer is the use of the technique of laser tweezers to form an optical trap, in which one end of the biopolymer can be held stationary [7]. If a biopolymer is emersed in a solvent environment, one can study flow-induced stretching as the background fluid starts to flow uniformly (Fig. 1c).

By now, we have encountered three basic types of mechanical manipulation of biopolymers. The observable that can be deduced from these experiments is usually displayed in an external strain versus molecular-extension curve, where the external strain can be a force, flow velocity, or a flow rate, depending on the design of the experiment. Other possibilities include displaying the extension–time or external strain–time curves. There is nothing extraordinary about the general shape of these curves; one pulls an object with a force and the object is stretched. One of the remarkable aspects of these curves, however, is that they often contain fine details that are believed to have a connection with certain physical processes that occur during stretching. Often a plateau or a zigzag in the strain–extension curve is observed. Two levels of interpretation to explain this are needed: at a level that gives a qualitatively microscopic explanation of the stretching kinetics and at a level that is correlated to the basic parameters in the bonding and conformational properties of the studied biopolymer. The latter is more difficult in view of the fact that an exact relationship between the bonding energy and the structure of a biopolymer, even for an unperturbed native state, is not straightforward. In addition, the significance of studying the behavior of a biopolymer in an elongational or uniform flow goes beyond external mechanical manipulation; the transport of DNA and protein in biological systems is associated with flow patterns that can be far more complicated than uniform or elongational flows, which are considered in most model systems.

The physical properties of proteins can be understood by studying statistical-physics models of polymers that capture the essential characteristics of real systems [8,9]. In such a coarse-grained model, the degrees of freedom of the coordinates involved are significantly reduced, in comparison with that of the all-atom modeling approach, which allows for the accumulation of adequate statistics in computer simulations. These types of model have been successfully used to explore the underlying physical mechanisms of protein structure formation, protein folding dynamics, biopolymer–biopolymer interaction, and the unfolding of protein and DNA under an external stretching force.

In the next two sections, a brief introduction to the basic concepts is given, aiming at introducing the general reader with a physics background to numerical simulation techniques suitable for studying biopolymers. Following the brief introduction, a more elaborative discussion is given in Section IV on the particular numerical technique used in a collisional molecular dynamics

(MD) simulation. Sections V–IX contain some examples of physical properties that can be examined by the use of MD techniques for the protein stretching simulation.

II. PROTEIN INTERMEDIATE STRUCTURES AND UNFOLDING EXPERIMENTS

The basic backbone of a protein molecule is a linear polymer chain that consists of repeated basic units, each containing an amide group NH, an α carbon Cα that connects to a side group, and a carbonyl group CO.* The bonds connecting N and CO of two different units form the basic peptide bond that forms during protein polymerization. There are 20 distinct side groups that can occur in proteins of common interest, thus yielding 20 different residues that act as the basic building blocks in proteins [10]. The primary structure simply states how these residues are connected linearly in a string of symbols in three-letter or one-letter format, each three-letter or one-letter label representing the name of an amino acid.

At room temperature, most protein chains present in an organism are organized into three-dimensional structures. The biological and physical functions of a protein are defined by the three-dimensional structure, which can be stabilized by the intraprotein and protein–solvent interactions and sometimes interprotein interactions in more involved multimolecular configurations. The one-dimensional primary sequence gives a prescription for a three-dimensional structure. With only a few exceptions, a unique protein sequence determines a unique three-dimensional structure [11].

In recent years, a wide range of statistical-physics models, in which various degrees of simplification to the real systems have been made according to the physical problem at hand, have established the predominant features of protein folding [12–24]. A key feature, critically important to the efficiency of the folding process, is the cooperativity of the potential energies of the entire system arranged in such a way that a protein has a funnellike free-energy landscape (Fig. 2) directed toward a native structure [16,19,20,23], the native state can then be found relatively easily by a random search in configuration space. Nucleation of a core structure is the basis of this folding concept, where the folding to the native structure builds upon the continuous growth

* A proline residue has a unique structure that is somewhat different from the reminder (see Ref. 10).

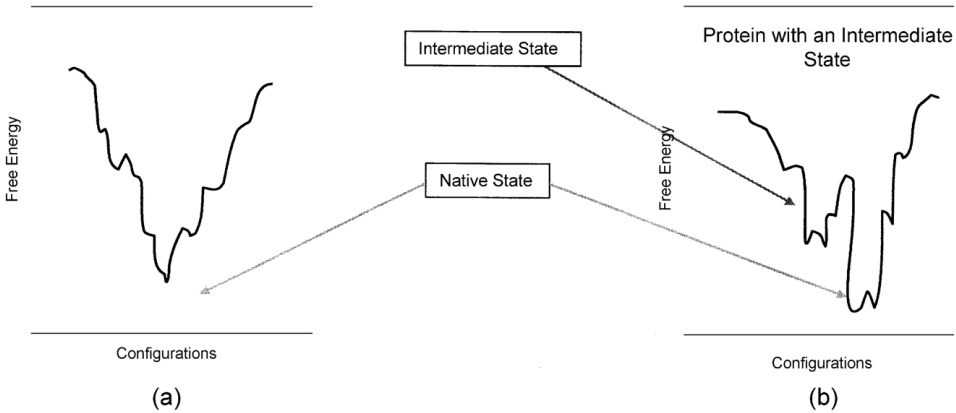


Figure 2 Illustration of the funnel structure of a protein's free-energy landscape: (a) a simple landscape where a protein can fold into the native state directly; (b) a more complicated landscape where an intermediate state exists that displays a relatively stable structure when a protein folds.

of a nucleated substructure [23,25–27]. The folding to the native structure can either be direct (Fig. 2a) or experience an intermediate state where one part of the protein is folded first with a relatively stable structure (Fig. 2b). Hence, the identification of possible intermediate states in the folding funnel in real proteins become critical for the verification of this basic folding physical picture.

Due to the existence of the secondary and tertiary structures of proteins [10], which are characteristically different from those of DNA, mechanical manipulation of proteins could demonstrate even richer physical phenomena than that of DNA. In a native state, protein molecules are in a compact conformation where the internal attraction, in particular the hydrophobicity and hydrogen-bonding interaction, plays a major role in forming the secondary and tertiary structures. Hence, the expectation is that the response of a protein to an external force field might reflect the bonding interaction and possibly directly show the folding intermediate states. Although potentially highly informative, direct experimentation has rarely been performed on protein denaturation in strong stretching flows, with the exception of the the uniform flow experiment [7]. Other forms of mechanical manipulations of proteins, in particular AFM experiments of the protein titin, have recently been reported [5]. In an effort to demonstrate the relevance of forced unfolding to the folding pathway, Carrion-Vazquez and co-workers have shown that the un-

folding rates of the 127 domain of the protein titin obtained by mechanical and chemical denaturation are the same and the transition state for the unfolding appears at the same position on the folding pathway [5].

III. COMPUTER SIMULATIONS OF STRETCHING A PROTEIN

Computer simulation studies of force-induced unfolding of proteins provide complementary information on the underlying microscopic physical picture of the denaturing process. Both types of model, all atom and coarse grained, are useful for this purpose [28–41]. For example, Klimov and Thirumalai have used a lattice model [38] and later more realistic β-bundle models [39] to study the unfolding intermediate stages when a protein is pulled by a force at both ends. The latter study indicated that thermal unfolding proceeds in all-or-none fashion, without any detectible intermediates. In contrast, force-induced unfolding of one of these models yields an observable intermediate. Paci and Karplus [31,34] have shown that the intermediate states detected in the forced unfolding simulations of all-atom protein models are not evidently related to those observed in a thermal denaturing experiment as the temperature increases. Whereas the studies of Paci and Karplus were carried out for systems containing no explicit water molecules, other groups have considered the presence of water molecules explicitly [28–30,33,37]. Lu et al. [28,29,33], for example, used steered MD simulations to predict the atomistic events during the forced extension of titin 127 domain. These simulations were successful in reproducing a force peak observed on the force–extension curve from AFM experiments. These simulations identified structural components responsible for a large bottleneck to unfolding of the 127 domain. A "key event" initiating domain unfolding consists of the simultaneous and cooperative disassociation of hydrogen bonds which connect the two β-strands on the C- and N-termini to the other part of the protein. We have investigated the behavior of a β-sheet barrel model in uniform and elongation flows [40] and found different unfolding scenarios in these two types of flow. Best et al. [37] have studied unfolding of two proteins, barnase and 127 domain, and concluded that the force-induced intermediate states found in computer simulations of pulling are much more nativelike than those found in simulations of thermal denaturation.

Hence, an important issue here is to what extent the intermediate states participate in these unfolding and folding events and whether they are related to refolding pathways. In order to give a more systematic comparison of the unfolding pathways experienced by proteins in a wider variety of unfolding environments, we have recently discussed the unfolding of two minimal pro-

tein models, an α-helix bundle [42] and a β-sheet barrel [39] under different types of denaturing environment: during uniform temperature increase, abrupt temperature jumps, in a uniform flow, and in an elongation flow [41]. Our simulation results have indicated that the unfolding pathways are critically dependent on the physical mechanics of how a protein is denatured. The relationship of forced unfolding pathways to the actual folding pathways is *not* always transparent.

The choice of these two simple potential models enables performing the extensive sampling of conformational space needed for an accurate calculation of the thermodynamical (heat capacity curves, phase diagrams, free-energy profiles) and kinetic (folding) properties of the system, which would otherwise be unfeasible in an all-atom approach. The richness of the resulting structures allows us to test structures containing two commonly occurring structures, α-helix and β-sheet, under different perturbing environments and to study the relationship between the forced-induced and refolding intermediate states. In particular, one additional advantage of using these two models is that their structural properties have been previously well documented [27,42–46]. The native structures, corresponding to the minima of the potential energy, are illustrated on the left top of Figs. 3a and 3b respectively for the α-helix bundle and β-barrel. The α-helix bundle has 73 monomers and the β-barrel has 46 monomers. Four segments of stable secondary structures are formed in each of these two models. The key components of the potential energy in both polymer models include terms that deal with the fluctuations of the bond angles about a fixed value, a heterogeneous long-range Lennard–Jones potential reflecting the three different types of residue (hydrophobic, hydrophilic, and neutral), and a torsional-angle potential that effectively prefers a local α-helix or β-sheet conformation [27,43]. Figure 3 demonstrates a series of representative structures as the temperature is lowered, together with the heat capacity and radius-of-gyration measurements that give signatures of the structural stability. The plots in Fig. 3 are based on our independent MD calculations by using the WHAM method [47]. In both models, there exists an intermediate state at a reduced temperature about $\widetilde{T} \equiv Tk_B/\epsilon \approx 0.6$, where the protein structure is partially bundled; ϵ is a typical energy scale when two residues are bonded through a hydrophobic force. In the α-helix bundle case, the intermediate state has a stable three helix bundle connected to a separated segment of a well-formed helix. The minimum in the heat capacity between two peaks signifies the existence and stability of this intermediate. In the β-barrel case, the intermediate state has a less stable three β-strand barrel connected to a separate segment. The segment displays coiled conformation and would form a β-sheet if bonded to the other structure. This intermediate state leaves no clear signature in the heat capacity plot and exists only as a kinetic folding intermediate as discussed below.

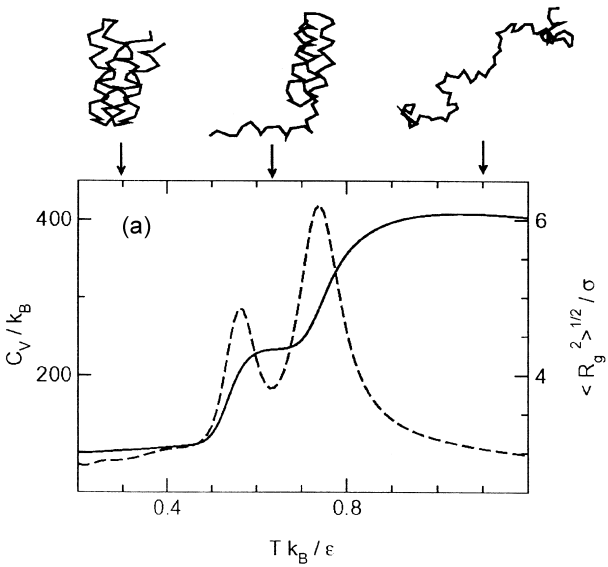


Figure 3 Temperature dependence of the heat capacity C_v (dashed line to the left scale) and the average radius of gyration $\langle R_g^2 \rangle^{1/2}$ (solid line to the right scale) for the α-bundle (a) and β-barrel (b) protein models. A schematic representation of the protein structures corresponding to different thermodynamic states is displayed on the top of each figure. The native states of the models, displayed at left of the plots, are a four-α-helix bundle and a four-β-strand barrel. The intermediate state in the middle of (a) is thermodynamically stable, as can be identified by the minimum in C_v. However, the intermediate state in the middle of (b) is thermodynamically less stable and leaves no traceable signatures in the C_v plot in (b). The latter is a kinetic intermediate (see text). (After Ref. 41.)

IV. MOLECULAR DYNAMICS METHOD

In this section, we discuss the main steps used in constructing the collisional MD simulation, suitable for studying the unfolding a protein in various types of velocity field. The central task in MD is to study the evolution of the time trajectory of particles in the system according to Newton's second law. Focusing on a selected particle, one has

$$m \frac{d^2\mathbf{r}(t)}{dt^2} = \mathbf{F} \tag{1}$$

where $\mathbf{r}$ is the coordinate of the particle of mass m, t is time, and $\mathbf{F}$ is the force acting on the particle which is a function of the coordinates of neighboring, if not all, particles in the system [48].

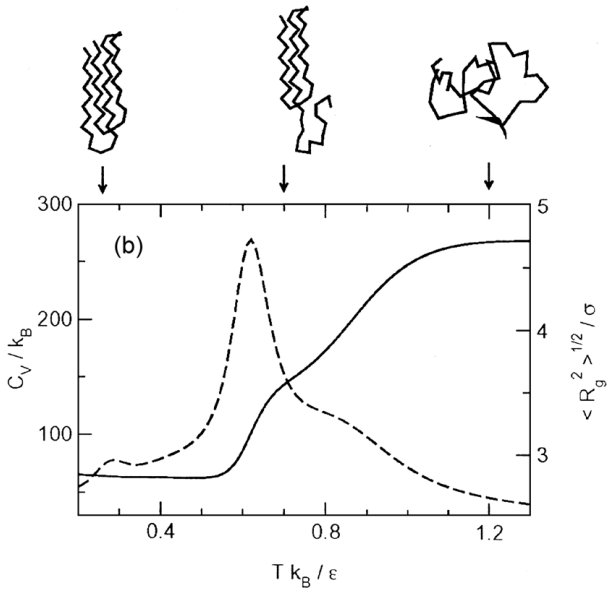


Figure 3 Continued.

In a given potential model, such as the α-helix bundle and β-sheet barrel models mentioned earlier, the functional dependence of **F** on the coordinates of the particles are specified. We are then required to solve this second-order ordinary differential equation numerically. The so-called "molecular dynamics simulation" is a numerical scheme that solves Newton's equation of a molecular system. Corresponding to a long microscopic period of time, Newton's equations are solved repeatedly to produce trajectories for the particles in the system, corresponding to a finite period of real-time observation.

After the replacement of the second-order derivative on the left-hand side of the force equation [Eq. (1)] with a finite difference, we can incorporate *Verlet's algorithm* for solving Newton's equation:

$$\mathbf{r}(t+h) = 2\mathbf{r}(t) - \mathbf{r}(t-h) + \mathbf{F}h^2 \tag{2}$$

where h is a time step and **F** is evaluated according to the coordinates of the particles at time t. The implementation of Verlet's algorithm in MD simulation is straightforward; the total energy, rather than temperature, is kept constant in such a simulation, provided that the accumulated error does not diverge.

In our simulations, the temperature of the system is effectively kept constant by incorporating interaction between the model proteins with the back-

ground solvent, in terms of collisional dynamics [49,50]. Monomers collide with virtual solvent particles that obey a Maxwell–Boltzmann velocity distribution. These stochastic collision events occur folloing a Poisson process which is specified by a single parameter, the collision frequency. Between the stochastic collision events, the dynamic system evolves as in a typical MD simulation according to Verlet's algorithm. The result of the collision event is an instantaneous change in the velocities of the monomers. Postcollision velocities are calculated by solving the collision problem [49], in which the velocity $\mathbf{v}_0$ of a virtual solvent particle is initially drawn randomly according to the distribution.

$$P(\mathbf{v}_0, \mathbf{r}) = \left(\frac{2\pi k_B T}{m_0}\right)^{-3/2} \exp\left(\frac{-m_0[\mathbf{v}_0 - \mathbf{w}(\mathbf{r})]^2}{2k_B T}\right) \tag{3}$$

where T is the solvent temperature and m_0 is the mass of the virtual solvent particle. The hydrodynamic velocity of the solvent at position $\mathbf{r}$, $\mathbf{w}(\mathbf{r})$, is required to simulate the flow environment and determines the type of flow force field. For simplicity, we also assume that the mass m of all residues involved in the system is the same, although we are dealing with a heteropolymer model in which different types of residue should, in principle, have a different mass.

To establish the simulation of a steady elongational flow of the virtual particles with a constant-velocity gradient, as suggested by the experiment of Perkins et al. [7], the velocity field can be selected to have the form

$$w_x = \dot{\epsilon}(x - x_c), \qquad w_y = -0.5\dot{\epsilon}(y - y_c), \qquad w_z = -0.5\dot{\epsilon}(z - z_c) \tag{4}$$

where $\dot{\epsilon}$ is the elongational rate. The fluid flows into the observation area centered at x_c, y_c, z_c in both y and z directions, and flows out in the x direction, where the center of mass, x_c, y_c, z_c, of the protein is assumed to move together with the flow pattern.

To establish the simulation of a steady, uniform flow, the velocity field does not depend on spatial coordinates and can be assumed in a simple form:

$$w_x = V, \qquad w_y = 0, \qquad w_z = 0 \tag{5}$$

where V is the flow velocity along the x direction. One of the two terminal monomers of the protein chain is assumed fixed. This velocity field effectively simulates the physical environment of a protein gripped by laser tweezers or an AFM tip at one end.

In these two setups, $\dot{\epsilon}$ and V are two essential physical parameters characterizing the external stretching strain on the protein system. For each fixed $\dot{\epsilon}$ or V, 30–40 MD runs, each consisting of 2×10^7 MD steps, have been performed in the simulations described in Section V. The time step h used in the simulation was set to $0.005t_0$, with $t_0 = \sigma(m/\epsilon)^{1/2}$ being a basic reduced

time unit and σ the bond length between the adjacent monomers. The collision frequency per monomer was chosen to be $1.2t_0^{-1}$.

V. β-BARREL MODEL IN AN ELONGATIONAL FLOW

In this section we concentrate on the demonstration of physical properties that can be calculated in these simulations by using the β-sheet barrel model as an example [44]. The temperature of the protein system is indirectly fixed by letting $T = T_L = 0.28\epsilon_0/k_B$ in Eq. (3). The folding temperature of this protein model to a four-stranded bundle structure is $T_f = 0.3\epsilon_0/k_B$ [44], above T_L. At T_L, the protein is in the native state and adopt a compact conformation in an equilibrium fluid. The hydrophobic attraction between monomers dominates, effectively producing a poor solvent environment in terms of polymer solution. We started the simulation by adopting the native structure, oriented randomly, as the initial configuration before we applied the flow to the system.

The solid circles in Fig. 4a demonstrate the dependence of average radius of gyration $\langle R_g^2 \rangle^{1/2}$ as a function of the elongation rate $\dot{\epsilon}$ defined in Eq. (4). Because of strong internal attractions among monomers in the protein, deformation does not begin until a rather large elongation rate is applied to the system. The passage from the native state to the stretched state is very sharp and abrupt. The root-mean-square (rms) fluctuations of the radius of gyration shown in Fig. 4b also indicate very large conformation fluctuations associated with this transition. No visible, metastable intermediate states were observed during the unfolding simulations.

At the same T_L, we have also investigated how the modelled protein responds to flow using the initial condition of a fully stretched state, to simulate the refolding process. The open circles in Fig 4a represent $\langle R_g^2 \rangle^{1/2}$ of the steady state as a function of the elongation rate $\dot{\epsilon}$. We see that the protein remains in the stretched state at intermediate flow rates and the collapsing to the native structure occurs only when a rather small flow rate is applied to the initial stretched state. This transition elongation rate is two orders of magnitude smaller than that required for the unfolding simulation. Again, the transition, as represented by the opened circles in Figs. 4a and 4b, is abrupt.

Figure 4c shows the distribution function of radius of gyration observed at different flow rates in the stretching simulations. At a relatively low elongational rate in the critical region, the $P(R_g)$ curve displays a sharp peak centered at a R_g value which corresponds to the compact globule state, and a second sharp peak, centered at a large R_g value which corresponds to the fully stretched state. The relative heights of these two peaks change as $\dot{\epsilon}$ is increased. At the critical elongation rate, the two peaks have almost the same

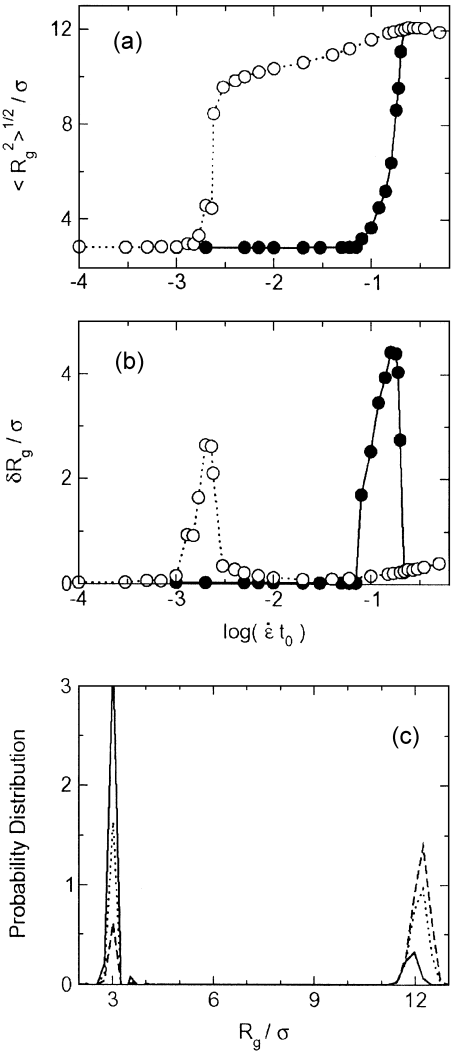


Figure 4 The average radius of gyration $\langle R_g^2 \rangle^{1/2}$ (a) and its root-mean-square deviation δR_g (b) versus elongation rate $\dot{\epsilon}$ for the β-barrel protein model in an elongational flow at $T_L = 0.28\epsilon_0/k_B$. The flow rate $\dot{\epsilon}$ is measured in terms of the basic reduced time unit t_0. Each data point represents the averaging performed over a number of different trajectories of observation time $t_{\text{obsrv}} = 10^5 t_0$. Initial configurations adopted in the simulations were highly stretched configurations (open circles) or native structures (filled circles). (c) Probability distribution of radius of gyration of the protein model in an elongational flow at $T_L = 0.28\epsilon_0/k_B$ and $\dot{\epsilon} = 0.12 t_0^{-1}$ (solid line), $0.18 t_0^{-1}$ (dotted line), and $0.2 t_0^{-1}$ (dashed line). The data were collected after a steady state has been established and reflect the fluctuations of the conformations about the mean value shown in (a). (After Ref. 40.)

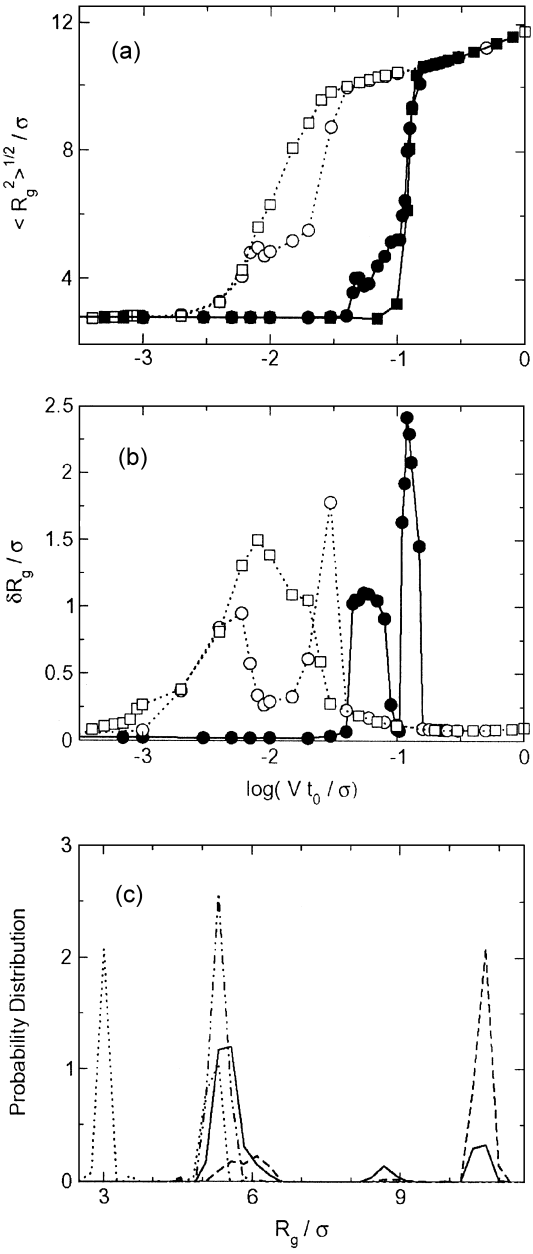
height. Finally, at a larger elongation rate, the distribution function $P(R_g)$ becomes singly peaked at a large R_g.

The above-described data indicate that the globule-to-stretch transition of the protein in the poor solvent condition shows the general feature of a strong first-order phase transition. If we design an experiment that gradually stretches protein from the native state, the native-to-stretch transition would take place at a greater $\dot{\epsilon}$ value than that of the reverse experiment in which we start from the stretched state and gradually reduce the elongation rate. A principal property of a first-order phase transition is the hysteresis effect [51] when the key adjustable parameter, in our case the elongation rate, first increases and then decreases. In analogy with the behavior of systems undergoing a first-order thermodynamic phase transition, there is a certain range of the elongation rates (a transition region) where the effective potential $W(R_g) = -k_B T \ln P(R_g)$ has two minima separated by a high barrier. In the transition region, the protein may exist in two states: the native and the fully stretched. Both are stable with respect to the thermal fluctuations. As noted previously [52–56], in a polymer system the hysteresis effect is inherently a time-dependent phenomenon that cannot be observed in *truly* steady-state flow conditions.

VI. β-BARREL MODEL IN A UNIFORM FLOW

In this section, we discuss the computer simulation of the case where one terminal end of the protein is tethered in space to a fixed point while the entire chain is subject to a uniform stretching flow. The β-sheet barrel model is represented by a linear heteropolymer chain of 46 monomers jointed together with bonds of length σ. Each monomer represents a pseudoresidue and can be one of three types: hydrophobic (H), hydrophilic (P), or neutral (N). The sequence specifying the β-barrel model is $H_9N_3(PH)_4N_3H_9N_3(PH)_5P$ [42].

First, we consider the case where the hydrophilic end (the last monomer) is tethered in space and observe the unfolding simulations with initial configurations adopted from a native structure. The filled circles in Fig. 5a represent the rms of the radius of gyration as a function of flow rate V. The existence of two intermediate transitions is clearly visible in Fig. 5a, where a rapid increase in $\langle R_g^2 \rangle^{1/2}$ is interrupted by a plateau in the transition region. The conformational fluctuations near the intermediate transitions are very large and can be viewed in Fig. 5b. Between the transition regions and near $V = 0.1\sigma t_0^{-1}$, there is a narrow range of flow velocities where the fluctuations of $\langle R_g^2 \rangle^{1/2}$ are considerably smaller than in the transition regions, indicating a well-defined, partially deformed structure in addition to the native ($\langle R_g^2 \rangle^{1/2} \approx 3.0\sigma$) and fully stretched ($\langle R_g^2 \rangle^{1/2} \approx 10.0\sigma$) states. Indeed, the distribution function $P(R_g)$ at this value of V has the form of a sharp single peak centered



at $R_g = 5.5\sigma$ (see the dot-dashed line in Fig. 5c). This is a stable stretching intermediate state.

The distribution functions of R_g at a few other values of V are shown in Fig. 5c. Depending on V, the $P(R_g)$ curve may display one, two, or three different peaks at one of the four specific values of R_g, namely at $R_g = 3$, 5.5, 8.5, and 10.5σ. This indicates the existence of four specific conformational states which are relatively stable and can be clearly observed over a relatively long time. Observation of the underlying microscopic behavior along individual MD trajectories allowed us to follow the conformational changes. The unfolding process could be broken down into three steps. In the native state, the four β-strands are packed together, forming an ordered compact globule state (see Fig. 6a,b). During the first stage of unfolding, the β-strand labeled D with mixed hydrophobic and hydrophilic residues is unzipped abruptly from the other three β-strands, resulting in a structure consisting of an isolated stretched strand and a three β-stranded bundle formed by two entirely hydrophobic A and C and one mixed β-strand B (see Fig. 6a). This stable structure has $R_g = 5.5\sigma$ and corresponds to the intermediate state observed in the $R_g(V)$ curve. At the next step of the unfolding process, the β-strand C detaches from the bundle part of the protein. The second intermediate structure consists of two stretched segments, C and D, and a double stranded β-sheet formed by A and B (see Fig. 6a). This structure has $R_g = 8.5\sigma$, corresponds to the kinetic intermediate of the unfolding process, and is relatively unstable. Only half of the residues involved in segment B are hydrophobic and can be attracted to the fully hydrophobic segment A. This relatively unstable structure leaves no trace in the R_g–V characteristics displayed in Fig. 5.

In the complementary refolding simulations, the initial configurations in the folding experiments were adopted from a fully stretch conformation and then the protein is subject to different flow velocities. Open circles in Fig. 5

Figure 5 The average radius of gyration $\langle R_g^2 \rangle^{1/2}$ (a) and its rms deviation δR_g (b) versus flow velocity V for the β-barrel protein model tethered by the hydrophobic (squares) and hydrophilic (circles) terminal ends in uniform flow, at $T_L = 0.28\epsilon_0/k_B$. Each data point represents the average performed along a number of different trajectories of observation time $t_{\text{obsrv}} = 10^5 t_0$. Initial configurations were selected from configurations in the stretched state (open circles and squares) and native state (filled circles and squares). Intermediate states in the curves associated with circles are clearly visible. A hysteresis loop is formed when a protein is stretched from a low flow rate to a high flow rate and folds back to a low flow rate. (c) Probability distribution of radius of gyration of the protein model tethered by the hydrophilic end in a uniform flow at $T_L = 0.28\epsilon_0/k_B$ and $V = 0.05\sigma t_0^{-1}$ (dotted line), $0.10\sigma t_0^{-1}$ (dot-dashed line), $0.12\sigma t_0^{-1}$ (solid line), and $0.14\sigma t_0^{-1}$ (dashed line). The distribution is significant in different regions where identifiable conformational states exist. (After Ref. 40.)

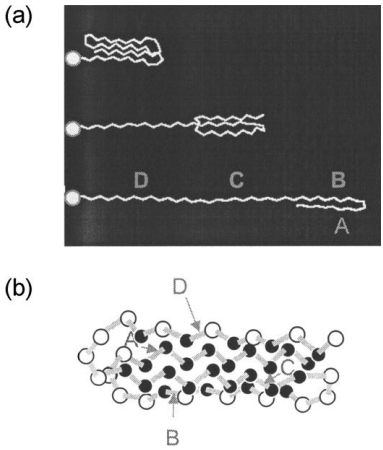


Figure 6 (a) Photograph of the intermediate configurations of the β-sheet barrel protein with the hydrophilic end tethered in a uniform flow (after Ref. 40) and (b) the native structure of the β-sheet barrel model with strands labeled (after Ref. 42). The three structures displayed in (a) are the unfolding steps observed in the simulations.

represent the R_g characteristics, which are different from the unfolding ones (solid circles). Again, the multistep folding process is clearly visible in these experiments. The low rms at $V = 0.01\sigma t_0^{-1}$ (Fig. 5b) is an indication of a relatively stable intermediate state, which has the same $R_g = 5.5\sigma$ as in the unfolding experiment. The time sequence of the folding events is similar to the plots displayed in Fig. 6, but in a reverse order (D bonds, C, B, and A sequentially).

Due to the sequence asymmetry along the chain, one can also tether the hydrophobic end (first monomer) of the protein to a fixed point in space. A very different unfolding (folding) scenario is then observed in the simulations. The solid (open) squares in Fig. 5 represent the R_g characteristics of the unfolding (folding) processes. No intermediate state could be seen on both $\langle R_g^2 \rangle^{1/2}$ and δR_g curves in Figs. 5a and 5b. The unfolding (folding) curves, however, display hysteresis. There are three possible conformational states in which the protein can reside and the unfolding process consists of two kinetic steps rather than three. During the first step, both β-strands A and B unzip from the bundle simultaneously, which yields a structure schematically similar to the second intermediate state illustrated in Fig. 6a, with reversed labeling ABCD. This conformation, however, is only short-live and the CD bonding is quickly detached in the second step.

The difference in the unfolding processes described here can be attributed to the architecture of the native structure of the β-barrel protein. The numbers

Table 1 Number of Tertiary Hydrophobic Contacts Among the Four β-Strands in the Native Structure of the β-Barrel Protein

	Strand B	Strand C	Strand D
Strand A	6	17	12
Strand B		10	0
Strand C			8

Note: Strands A, B, C, and D contain monomers 1–9, 13–20, 24–32, and 36–46, respectively.
Source: Ref. 40.

of hydrophobic contacts between different β-strands in the native structure are shown in Table 1. One can see that 20 hydrophobic contacts are broken when the mixed strand D is unzipped from the other three strands, whereas 35 hydrophobic contacts need to be broken in order to unzip the entirely hydrophobic strand A from the other three. Three strands (A, B and C) are kept intact by 33 contacts while strands B, C and D are fused together by only 18 contacts. Thus, the three-bundle ABC has a hydrophobic core which is almost twice as large as the hydrophobic core in the three-bundle BCD. This is the reason that the structure consisting of unzipped strand D and bundle ABC is a more stable intermediate in the unfolding process induced by uniform flow.

The same kind of intermediate structure was observed by Klimov and Thirumalai [39] in their recent study of forced stretching of two model sequences which have four-stranded β-barrel native topology. In their simulations, a model protein was stretched by pulling two terminal ends in the opposite directions simultaneously. In the native structure of one of the model sequences (sequence S2 in Ref. 39), strand A forms 1.6 times more contacts than that formed by strand D, which is similar to the interactions in the native structure of the β-barrel model studied here. It was also shown that sequence S2 unfolds in a two-step process going from the native state to the stretched state through a stable intermediate in which strand D is stretched and strands A, B, and C maintain their native conformation. No other kinetic intermediate structures have been observed in their forced stretching simulations.

VII. UNIFORM FLOW AND THE FORCE–EXTENSION CURVE

In this section, we discuss the similarity between the stretching of a protein under a uniform flow, where one terminal end of the protein is fixed in space (Fig. 7a), and pulling a protein by applying a mechanical unzipping force

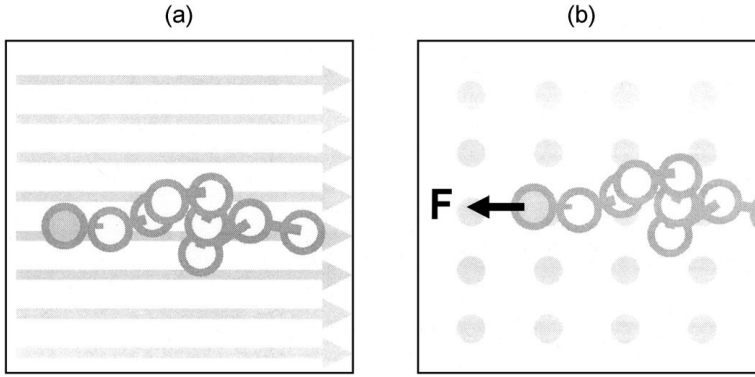


Figure 7 (a) A reference frame stationary to the tethered terminal end of the protein and (b) a reference frame moving together with the constant velocity field. In (b), the terminal end moves with constant velocity and appears to have been pulled by a force **F**, which varies because of both the changes in the protein conformation and the thermal fluctuations in the system.

on a terminal end, where the entire protein chain is in an equilibrium fluid (Fig. 7b). The two physical scenarios are connected by the choice of the reference frame; an inertia moving frame will not change the physical properties. The latter perspective is more similar to the experimental studies of mechanical unzipping of two proteins with a β-sandwich architecture (Ig27 domain of the protein titin [4] and FnIII domain of tenascin [6]) by AFM. One common method to illustrate these experimental data is to present a force versus molecular extension diagram.

Examining the force acting on the fixed terminal end of the protein chain, we can recreate such a force–extension distribution as illustrated in Fig. 8. Because of the thermal fluctuations, the effective force **F** acting on the terminal end has a distribution of values. In Fig. 8, R_x represents the dimension of the model protein in the x direction to which the force is applied. As a projection of the distribution function $\rho(\mathbf{F}, R_x)$ onto the $\mathbf{F}$–R_x plane, the darker color in Fig. 8 presents a higher probability that a particular pair of **F** and R_x might appear, in terms of histogram collected in the simulations. There are four distinct regions where the **F** and R_x values are preferentially distributed, which indicate stepwise unzipping of the protein under different flow rates.

For a fixed R_x, there is a probability distribution of force acting on the terminal end, with the average value $\langle \mathbf{F} \rangle$ shown in Fig. 8 as circles. The stretching of a protein starting from its native state can be characterized by the force–extension $\langle \mathbf{F} \rangle$ curve. Initially, the intramolecular energy interac-

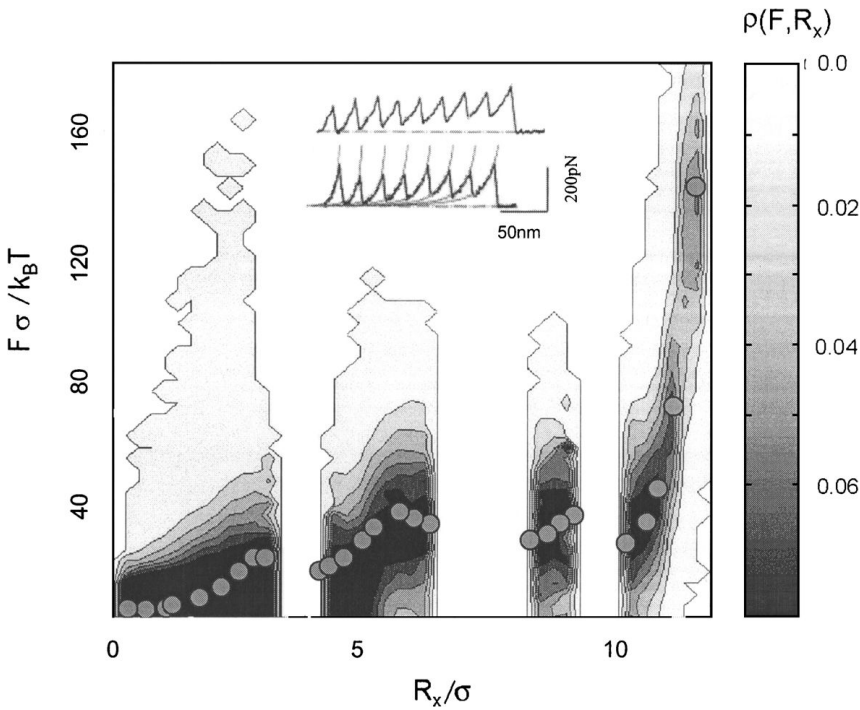


Figure 8 Force–extension probability distribution $\rho(\mathbf{F}, R_x)$ for the β-barrel protein tethered by the hydrophilic terminal end in the uniform flow in an unfolding experiment at temperature $T = 0.28\epsilon_0/k_B$. Contour surfaces have been plotted here. The probability distribution scale is gray-scale coded as shown on the scale bar. Four distinct states are visible in this plot. The filled circles represent the mean force acting on the anchored end (hydrophilic) of the protein versus the extension in the direction of the uniform flow (the x direction). The inset shows the force–extension curve for the protein titin obtained in AFM experiment. Each peak on the curve correspond to the unfolding of a single Ig domain. (After Ref. 40.)

tions dominate and resist the potential energy imposed by the pulling force. There are two clearly visible unzipping stages where the force decreases after unzipping a particular bonded structure: One occurs in the region $3\sigma < R_x < 4.5\sigma$ and the other in $5.5\sigma < R_x < 10\sigma$. This is a signature of two separate configurational transitions associated with the two abrupt unzipping events in the protein unfolding process described earlier. The protein structures between successive unzipping events are composed of partially native and partially stretcheds (see Fig. 6). When the protein is in the initial native state or the intermediate partially deformed states, we observed an almost linear

increase of R_x with the increased F_x. An unzipping event occurs abruptly at a critical force that leads to a large increase in **F**. Shown in the inset is the force–extension curve of unfolding the protein titin performed by the AFM experiment, where a similar saw-shaped behavior can be understood as the stepwise unfolding of domains. The same feature of hysteresis discussed in the last section was also observed in experimental studies of the unfolding of single molecules of the protein titin [4,57,58].

Klimov and Thirumalai used a 36-mer lattice model in which the contact energies are distributed according to the random energy model [38]. When the polymer is pulled *on both ends*, they also found that intermediate states exist depending on the pulling force. Lu and Schulten recently used the steered MD simulation to examinine force-induced unfoling of the 127 domain of the protein titin, with special attention paid to folding intermediates [29]. As an observed intermediate state, a structure with the hydrogen bonds between two β-strands, labeled A and G in their study, are broken, whereas hydrogen bonds between other β-strands are kept almost intact. There is no direct simulation of hydrogen bonds in the model used here, and the formation of β-sheets relies on a combined dihedral and hydrophobic potential. Nevertheless, the unzipping of only two β-strands as the first unfolding event described here qualitatively agrees with Lu and Schulten's observation.

VIII. THERMAL HEATING AND FLOW STRETCHING SIMULATIONS

In this section, we discuss the relationship between the characteristics of the physical properties observed in thermal heating and flow stretching simulations.

To simulate the thermal denaturing process in which the system is slowly heated up, we assume that temperature $\widetilde{T} = Tk_B/\epsilon$ is increased at a constant rate continuously from $\widetilde{T}_0 = 0.1$ to 1.0 in the α-bundle model and from $\widetilde{T}_0 = 0.1$ to 1.2 in the β-barrel model:

$$\widetilde{T}(\tilde{t}) = \widetilde{T}_0 + v_T\tilde{t} \tag{6}$$

where $\tilde{t}$ is the reduced time measured in terms of the MD steps. At initial temperature $\widetilde{T}_0$, the model proteins adopt the thermodynamically stable, native conformations. The final temperature is sufficiently high so that the model proteins can be clearly seen to adopt a random coil conformation (see the right top sketch in Fig. 3a and Fig. 3b). The solid and dashed curves in Figs. 9a and 9b demonstrate the heating $\overline{E}$–R_g characteristics of the two systems where $\overline{E} = E/\epsilon$ is the energy of the system. For the α-bundle model, heating rates of $v_T = 4 \times 10^{-8}$ and 9×10^{-8} were used, and for the β-barrel

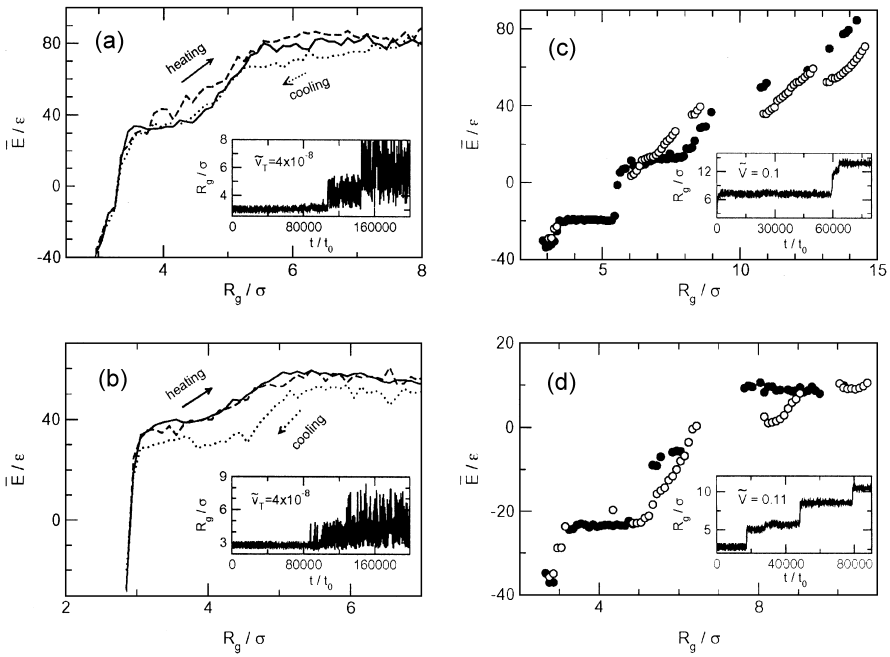


Figure 9 Statistical average energy $\overline{E}$ versus radius of gyration R_g calculated from three different MD trajectories of the α-bundle (a) and β-barrel (b) protein models in thermal heating/cooling simulations, and the α-bundle (c) and β-barrel (d) protein models in the uniform-flow simulations with different velocity values: (c) $\tilde{V} = 0.1$ (open circles) and 0.0007 (filled circles); (d) $\tilde{V} = 0.11$ (open circles) and 0.001 (filled circles). A native initial configuration is adopted in unfolding simulations (open circles) and a stretched initial configuration is adopted in refolding simulations (filled circles). The insets in (c) and (d) show the radius of gyration as a function of time in unfolding simulations at high flow velocities. In all these simulations, a terminal end of the protein is fixed in space. In (a) and (b), solid and dashed lines correspond to the heating experiments in which the temperature of the system was increased continuously with a constant rate of $v_T = 4 \times 10^{-8}$ and $v_T = 9 \times 10^{-8}$ (a) and $v_T = 3 \times 10^{-8}$ and $v_T = 11 \times 10^{-8}$ (b). The dotted lines correspond to the cooling experiments in which the temperature was decreased continuously with a constant rate of $v_T = -4 \times 10^{-8}$. The insets show the time behavior of the radius of gyration in the heating experiments with $v_T = 4 \times 10^{-8}$.

model, $v_T = 3 \times 10^{-8}$ and 11×10^{-8}. These different heating rates, however, can be seen to produce no major difference in these plots. In both models, the $\overline{E}$–R_g characteristic curves are relatively smooth, with a hint of the stability of the intermediate state region near $R_g = 4\sigma$ (see insets in Fig. 9) when the temperature passes through $\widetilde{T} = 0.6$, a temperature corresponding to the stable region of such an intermediate structure. Because the intermediate state is less stable thermodynamically in the β-barrel model, a less clear step is shown in the inset of Fig 9b.

In order to observe the annealed refolding, we slowly cool down the system by linearly decreasing $\widetilde{T}$ continuously from $\widetilde{T}_0 = 1.0$ to $\widetilde{T} = 0.2$ in both models, with a cooling rate of $v_T = -4 \times 10^{-8}$ in Eq. (6). An random coil structure was used as the starting point. The $\overline{E}$–R_g characteristics are demonstrated in Fig. 9 by dotted curves. In the β-barrel model, a plateau in the region near $R_g = 4\sigma$ is more clearly visible, which gives a signature of the kinetic stability of the intermediate state; energetically, the systems are trapped in these regions before an even lower energy can be attained. Such a trap, however, does not exist in the heating procedure, where the protein can continuously unfold giving rise to the rather smooth characteristic curves.

In comparison, consider the case when the modeled proteins are stretched in a simulated uniform flow environment at $T = 0.1\epsilon/k_B$ with the position of one of the terminal ends fixed in space. We have chosen a velocity field of $\widetilde{V} \equiv V_x t_0 \sigma^{-1} = 0.1$ that is coupled to the velocity distribution of the virtual solvent particles. This flow velocity adequately covers the entire stretching regime where the protein unfolds through more than one intermediate steps (see Section VI). The open circles in Figs. 9c and 9d represent the unfolding $\overline{E}$–R_g characteristics, where a few plateaus can be seen in both plots. In comparison with the above thermal denaturing simulations, the $\overline{E}$–R_g characteristics of both models show more features that are typical to a forced stretching experiment. In the vicinity of $R_g = 4\sigma$, the system is denatured through an intermediate state described earlier. Because the system temperature is relatively low, the secondary structures remain almost intact until the final stretching stage. The steps occurring at large R_g are not seen in thermal denaturing simulations. The additional intermediate state contains a stretched segment corresponding to two original α-helical (or β-sheet) segments and a segment which still has a partial native conformation. The plateaus on $\overline{E}$–R_g curves reflect the stepwise unzipping of secondary structure segments from a native structure before a fully stretched conformation can be achieved. The plots in the insets of Fig. 9 show R_g as a function of simulation time, where the steps that R_g going through are clearly visible.

Also plotted as filled circles in Figs. 9c and 9d are the $\overline{E}$–R_g characteristics of a refolding experiment where we have taken the fully stretched, final structure from the above simulations and observed the refolding of the modeled

proteins back to their native structures in a flow with low velocity. Starting from the initial extended, straight-line-like configurations, proteins refold by going through various stages in an almost reverse order from the unfolding process.

IX. CONCLUDING REMARKS

Our first conclusion is that proteins go through different unfolding intermediate stages when subjected to a denaturing force, which are not necessarily the same as the thermodynamically stable intermediate state. One notable result is that these intermediate states are completely absent in the elongational-flow simulation and temperature-jump simulations [41]. The dominating physical mechanisms are similar in these two cases: Every monomer receives an average kinetic energy that is great enough to break all its native contacts.

The thermodynamical intermediate states are clearly visible in both low-rate heating and cooling simulations where every monomer receives a small amount of kinetic energy and the systems slowly evolve into the final stage. At each time step, the systems are in quasiequilibrium. One of the features in the cooling experiment is that the intermediate state is reached at a relatively low temperature and hydrophobic contacts in the intermediate state can energetically trap the protein. Hence, a clearer intermediate step in the characteristic curve is observed.

The computer simulations of the 127 domain using implicit and explicit solvent models [28,29,31,33] have indicated an unfolding "bottleneck" (i.e., an intermediate state). MD simulations of these all-atom models of the 127 domain also revealed an energy landscape with extra minima. A comparable situation exists here (see Fig 2).

A different perspective of the folding and refolding simulations in an uniform flow can be gained with a reference frame moving with the flow. The fixed end becomes a moving monomer pulled by a force. No other monomers directly experience the pulling force. The segments that are connected to the moving monomer experience an indirect force due to the pulling. Hence, the pulling force rips the protein apart segment by segment, with some native domains still kept intact. A protein in a uniform flow or a protein pulled by one end thus exhibits richer unzipping intermediate states than that of a protein undergoing a thermal denaturation.

For the two protein models we have studied, the differences in the characteristic plots are caused by the different denaturing processes, not by the differences in the internal structure of the proteins. The intermediate states of the two models have different thermodynamic properties: stable for α-helix bundle and less stable for the β-barrel. These intermediate states, however,

play a crucial role kinetically in the formation of the native structures [44]. Uniform flow and thermal denaturing simulations more clearly reveal the intermediate states.

ACKNOWLEDGMENT

The financial support of this work was provided by the Natural Science and Engineering Research Council of Canada. The computation time was partially provided by SHARCNET.

REFERENCES

1. Perkins, T.T.; Smith, D.E.; Chu, S. Science 1997, *276*, 2016–2021.
2. Smith, D.E.; Chu, S. Science 1998, *281*, 1335–1340.
3. Fisher, T.E.; Marszalek, P.E; Fernandez, J.M. Nature Struct. Biol. 2000, *7*, 719–724.
4. Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandes, J.M.; Gaub, H.E. Science 1997, *276*, 1109–1112.
5. Carrion-Vazquez, M.; Oberhauser, A.F.; Fowler, S.B.; Marszalek, P.E.; Broedel, S.E.; Clarke, J.; Fernandez, J.M. Proc. Natl. Acad. Sci. U.S.A. 1999, *96*, 3694–3699.
6. Oberhauser, A.F.; Marszalek, P.E.; Erickson, H.; Fernandez, J.M. Nature 1998, *393*, 181–185.
7. Perkins, T.T.; Smith, D.E.; Larson, R.G.; Chu, S. Science 1995, *268*, 83–87.
8. Eaton, W.A.; Muñoz, V.; Hagen, S.J.; Jas, G.S.; Lapidus, L.J.; Henry, E.R. Annu. Rev. Biophys. Biomol. Struct. 2000, *29*, 327–359.
9. Shea, J.-E.; Brooks, C.L. III Annu. Rev. Phys. Chem. 2001, *52*, 499–535.
10. Fersht, A.R. *Structure and Function in Protein Science*; W.H. Freeman: New York, 1999: Brander, C.; Tooze, J. *Introduction to Protein Structure*; Garland Publishing: New York, 1991.
11. See http://www.rcsb.org/pdb/.
12. Miyazawa, S.; Jernigan, R. Macromolecules 1985, *18*, 534.
13. Lau, K.F.; Dill, K.A. Macromolecules 1989, *22*, 3986.
14. Chan, H.S.; Dill, K.A. Phys. Today 1993, *46*, 24.
15. Šali, A.; Shakhnovich, E.I.; Karplus, M. Nature 1994, *369*, 248–251.
16. Wolynes, P.G.; Onuchic, J.N.; Thirumalai, D. Science 1995, *267*, 1619–1620.
17. Karplus, M.; Šali, A. Curr. Opin. Struct. Biol. 1995, *5*, 58–73.
18. Gutin, A.M.; Abkevich, V.I.; Shakhnovich, E.I. Phys. Rev. Lett. 1996, *77*, 5433–5436.
19. Šali, A.; Shakhnovich, E.I.; Karplus, M. J. Mol. Biol. 1996, *235*, 1614–1636.
20. Klimov, D.K.; Thirumalai, D. Phys. Rev. Lett. 1996, *76*, 4070–4073.
21. Li, H.; Helling, R.; Tang, C.; Wingreen, N. Science 1996, *273*, 666–669.

22. Li, H.; Tang, C.; Wingreen, N. Phys. Rev. Lett. 1997, *79*, 765–768.
23. Shakhnovich, E.I. Curr. Opin. Struct. Biol. 1997, *7*, 29–40.
24. Klimov, D.K.; Thirumalai, D. Phys. Rev. Lett. 1997, *79*, 317–320.
25. Ball, K.D.; Berry, R.S.; Kunz, R.E.; Li, F.; Proykava, A.; Wales, D.J. Science 1996, *271*, 963–966.
26. Shakhnovich, E.I.; Abkevich, V.; Ptitsyn, O. Nature 1996, *379*, 96–98.
27. Guo, Z.; Thirumalai, D. Biopolymers 1995, *36*, 83–102.
28. Lu, H.; Isralewitz, B.; Krammer, A.; Vogel, V.; Schulten, K. Biophys. J. 1998, *75*, 662–671.
29. Lu, H.; Schulten, K. Proteins 1999, *35*, 453–463.
30. Krammer, A.; Lu, H.; Isralewitz, B.; Schulten, K.; Vogel, V. Proc. Natl. Acad. Sci. U.S.A. 1999, *96*, 1351–1356.
31. Paci, E.; Karplus, M. J. Mol. Biol. 1999, *288*, 441–459.
32. Socci, N.D.; Onuchic, J.N.; Wolynes, P.G. Proc. Natl. Acad. Sci. U.S.A. 1999, *96*, 2031–3035.
33. Lu, H.; Schulten, K. Biophys. J. 2000, *79*, 51–65.
34. Paci, E.; Karplus, M. Proc. Natl. Acad. Sci. U.S.A. 2000, *97*, 6521–6526.
35. Bryant, Z.; Pande, V.S.; Rokhsar, D.S. Biophys. J. 2000, *78*, 584–589.
36. Craig, D.; Krammer, A.; Schulten, K.; Vogel, V. Proc. Natl. Acad. Sci. U.S.A. 2001, *98*, 5590–5595.
37. Best, R.B.; Li, B.; Steward, A.; Daggett, V.; Clarke, J. Biophys. J. 2001, *81*, 2344–2356.
38. Klimov, D.K.; Thirumalai, D. Proc. Natl. Acad. Sci. U.S.A. 1999, *96*, 6166–6170.
39. Klimov, D.K.; Thirumalai, D. Proc. Natl. Acad. Sci. U.S.A. 2000, *97*, 7254–7259.
40. Lemak, A.S.; Lepock, J.R.; Chen, J.Z.Y. Proteins 2003, *51*, 224–235.
41. Lemak, A.S.; Lepock, J.R.; Chen, J.Z.Y. Phys. Rev. E 2003, *67*, 031910.
42. Honeycutt, J.D.; Thirumalai, D. Biopolymers 1992, *32*, 695–709.
43. Guo, Z.; Thirumalai, D. J. Mol. Biol. 1996, *263*, 323–343.
44. Guo, Z.; Brooks, C.L. III Biopolymers 1997, *42*, 745–757.
45. Shea, J.-E.; Nochomovitz, Y.D.; Guo, Z.; Brooks, C.L. III J. Chem. Phys. 1998, *109*, 2895–2903.
46. Nymeyer, H.; Garcia, A.E.; Onuchic, J.N. Natl. Acad. Sci. U.S.A. 1998, *95*, 5921–5928.
47. Kumar, S.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A.; Rosenberg, J.M. J. Comput. Chem. 1992, *13*, 1011–1021.
48. Allen, M.; Tildesley, D. *Computer Simulations in Liquids*; Oxford Science Publications: Oxford, 1994.
49. Lemak, A.S.; Balabaev, N.K. J. Comput. Chem. 1996, *17*, 1685–1695.
50. Lemak, A.S.; Balabaev, N.K. Mol. Simulat. 1995, *15*, 223–231.
51. De Gennes, P.G. J. Chem. Phys. 1974, *60*, 5030–5042.
52. Nguyen, T.D.; Kausch, H.H. *Flexible Polymer Chain Dynamics in Elongational Flow: Theory and Experiment*; Springer-Verlag: Herdelberg, 1999.
53. Wiest, J.M.; Wedgewood, L.E.; Bird, R.B. J. Chem. Phys. 1989, *90*, 586–594.

54. Darinskii, A.A.; Saphiannikova, M.G. J. Non-Crystal Solids 1994, *172–174*, 932–934.
55. Agarwal, U.S.; Bhargava, R.; Mashelkar, R.A. J. Chem. Phys. 1998, *108*, 1610–1617.
56. Neumann, R.M. J. Chem. Phys. 1999, *110*, 7513–7515.
57. Kellermayer, M.S.Z.; Smith, S.B.; Granzier, H.L.; Bustamante, C. Science 1997, *276*, 1112–1116.
58. Tskhovrebova, L.; Trinick, J.; Sleep, J.A.; Simmons, R.M. Nature 1997, *387*, 308–312.

9

Structure–Function Relationships of Aspartic Proteinases

RICKEY Y. YADA and TAKUJI TANAKA* University of Guelph, Guelph, Ontario, Canada

I. INTRODUCTION

In this chapter, the authors discuss the research undertaken in our laboratory on the use of genetic engineering techniques to refine the structure–function relationship of food-related proteins (particularly aspartic proteinases) at a fundamental, molecular level. In this light, we have focused on four areas: the role of prosegment residues to the structure–function relationship of the zymogen and activated enzyme; the role of residues in the flap loop region to the structure–function relationship of the active enzyme; the role of various residues/regions and posttranslational modification to functional stability; and the universality of structure–function relationships among the aspartic proteinases.

The biological activity of a protein depends on its structure. Therefore, the ability to predict the structure and function of a protein (enzyme) is the ultimate goal of protein biochemists and enzymologists. However, the exact nature of this relationship has remained relatively elusive. The advent of genetic engineering techniques has greatly aided our ability to refine these relationships. Among its many benefits, these techniques have revolutionized protein science by allowing for rapid sequencing of proteins from their DNA and for the systematic alteration of protein structures by gene mutation (i.e., protein engineering). In this regard, protein engineering is of fundamental importance in analyzing structure–function relationships in proteins. Germane, therefore, to the optimal use of food enzymes is a fundamental understanding of structure–function relationships.

*_Current affiliation_: University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

In the food area, aspartic proteinases play a prominent role as processing aids in the manufacture of a variety of products (e.g., chymosin is used to coagulate milk in the production of cheese). Aspartic proteinases are characterized by two catalytic aspartate residues localized in two short amino acid stretches with sequence homology to other aspartic proteinases, similar three-dimensional structures, optimal catalytic activity at acidic pH values, a scission preference between large hydrophobic amino acids, and inhibition by the naturally occurring compound pepstatin [1]. Although aspartic proteinases generally conform to these characteristics, there are exceptions: for example, differences in pH stability, substrate specificity, glycosylated, and nonglycosylated. It is this diversity that permits these enzymes to be an excellent model for structure–function studies.

II. RECOMBINANT PEPSINOGEN AND INCLUSION BODIES

A critical step in structure–function studies which use genetic engineering techniques is the choice of an appropriate protein expression system. Many researchers have used prokaryotic systems such as *Escherichia coli*; however, the cloned protein is often expressed as an inclusion body. Inclusion-body proteins require that they be unfolded and then refolded in order to obtain a "properly" folded protein. Many research groups have reported that the folding step is protein dependent and that a successful method for one protein does not always apply to other proteins [2–4]. Such results suggest that slight differences between experiments might lead to different forms of protein (i.e. refolded and misfolded protein). In addition, the refolding procedure does not ensure that all of the protein molecules have the identical and correct structure. If the protein is expressed in the soluble fraction as an active protein and shows the identical specific activity to the native pepsin, it is assumed that every protein molecule is correctly folded. Soluble expression systems are, therefore, superior to the insoluble expression systems.

Fusion protein expression systems (i.e., where the protein of interest is fused to an endogenous protein from the expression organism) have been used to obtain soluble proteins [5,6]. For our research, porcine pepsinogen gene was fused with thioredoxin gene and expressed in *E. coli* in a soluble form [7]. Using this system, we were able to produce soluble pepsin and pepsinogen molecules which had properties (milk clotting, proteolysis, kinetic parameters, pH dependency) similar to that of commercial pepsin(ogen) (Table 1 and Fig. 1). The above results, therefore, indicate the fusion system successfully produced recombinant porcine pepsinogen as a properly folded protein, which can be activated into active pepsin. It was noted, however, that our constructed fusion protein (thioredoxin–pepsinogen) exhibited dominant

Table 1 Kinetic Analysis of Recombinant and Commercial Pepsin

Enzyme	Milk-clotting activity (units/mg)	Proteolytic activity (units/mg)	K_m (mM)	k_0 (s^{-1})
r-Pepsin[a]	27.9 ± 1.5	19.8 ± 0.3	0.033 ± 0.005	65.4 ± 3.1
r-Pepsin[b]	27.7 ± 1.6	ND	ND	ND
n-Pepsin	28.3 ± 0.8	21.1 ± 0.5	0.026 ± 0.004	79.5 ± 3.8

ND: not determined.
[a] r-Pepsin purified directly from Trx-PG.
[b] r-Pepsin purified from r-pepsinogen.
Source: Ref. 7.

self-activation (unimolecular, intramolecular), in which a pepsinogen molecule cleaves itself to yield a pepsin molecule [8–11] in contrast to pepsinogen which exhibited both unimolecular and bimolecular reactions (intermolecular activation in which a pepsin molecule released during activation converts pepsinogen into pepsin) and was attributed to the thioredoxin (Fig. 2) [12]. Only when thioredoxin was covalently bonded to pepsinogen was the unimolecular activation dominant. The thioredoxin portion is approximately one-quarter of the entire fusion molecule (Fig. 2). The bulky "prosegment" of the fusion protein would stabilize the catalytic intermediate of unimolecular

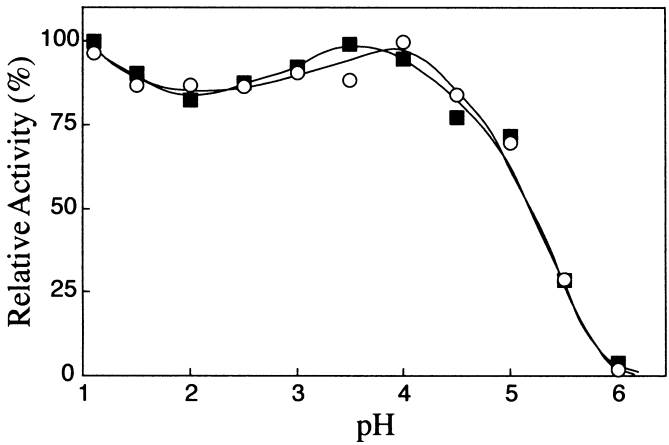
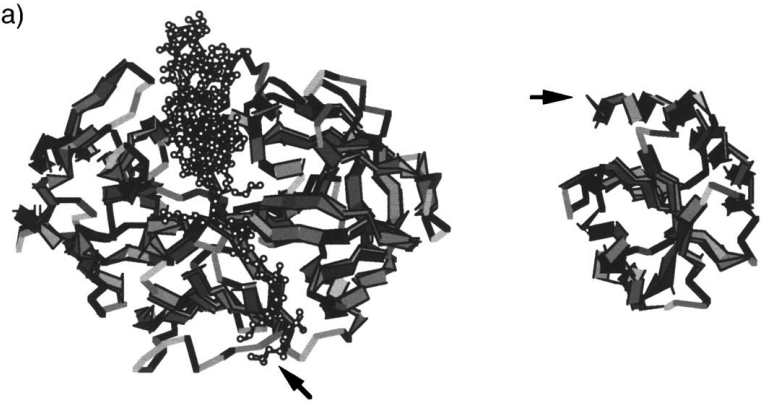
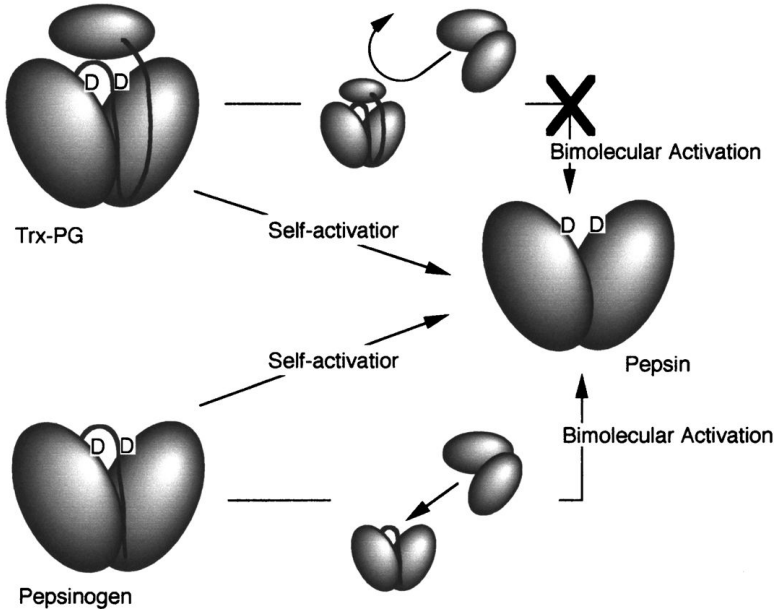


Figure 1 The pH dependency of hydrolysis of synthetic peptide substrate with pepsins. Hydrolysis of the synthetic substrate was measured at the approprite pH. Data are reported as percentage relative activity, a percentage of the highest activity over the pH range examined for the enzyme in question. Open circles and closed boxes represent r-pepsin and n-pepsin, respectively. (From Ref. 7.)

(a)



(b)



activation and could also retard bimolecular activation. The bulkiness of the prosegment of the fusion protein could prevent pepsin molecules from approaching the cleavage site, Leu44p-Ile1 (Fig. 2).

In addition to examining the role of specific amino acid residues/regions, our group was interested in the role of posttranslational modification (e.g., glycosylation) to the structure–function relationships of aspartic proteinases. Although fusion systems allow for the expression of soluble protein, procaryotic systems involving *E. coli* do not possess any posttranslational modification system. Therefore, in structure–function studies in which the effect of the addition of ligands, such as carbohydrates, is to be studied, an alternative expression system(s) must be used.

Recently, the methylotrophic yeast *Pichia pastoris* has become a dominant tool in molecular biology for the production of recombinant proteins. As an eukaryote, it is capable of posttranslational modifications during expression such as proteolytic processing, folding, disulfide bond formation, and glycosylation [13]. This system also has the advantage over intracellular expression systems because most of the protein in the culture media will be the target proteins, thus leading to an easier purification. In this light, our group had success in expressing pepsinogen extracellularly from *P. pastoris* [14]. Subsequent activation of the zymogen to pepsin indicated that the K_m and k_{cat} values for both commercial and recombinant pepsins (Table 2) were not significantly different ($p > 0.05$), therefore indicating similar enzymatic activity. In addition there were no differences in pH dependency of the activity (Fig. 3). These results indicated *Pichia* express cloned pepsinogen as soluble natural form protein.

Figure 2 Three-dimensional structure of pepsinogen and thioredoxin and schematic scheme of activation of Trx-PG. (a) Three-dimensional structures of pepsinogen and thioredoxin (Protein Data Bank 3PSG and 2TRX, respectively) are shown in the same scale. The prosegment of pepsinogen is shown in ball-and-stick models. The N-terminal of pepsinogen and C-terminal of thioredoxin are indicated with arrows. Thioredoxin consists of 108 amino cid residues, whereas pepsinogen is 371 amino acid residues long. In Trx-PG, both protein are connected by a 20-amino-acid residue linker. (b) Proposed scheme of how thioredoxin prevents pepsin from cleaving the fusion protein. In both the fusion protein (Trx-PG) and pepsinogen, two aspartic active sites (two D's in the figure) are covered with a prosegment (the thick lines in the figure). Pepsinogen can be activated into pepsin through either self-activation or bimolecular activation. Trx-PG has a large independent domain (i.e., thioredoxin portion) at the amino terminal of the prosegment. This large domain would prevent pepsin from approaching the susceptible site on the prosegment, and therefore, the bimolecular reaction could not occur. (From Ref. 12.)

Table 2 Kinetic Analysis of Recombinant and Commercial Pepsin

Enzyme[a]	pH 1.0	pH 2.0	pH 3.0
r-Pepsin			
K_m (mM)	0.091 ± 0.010	0.062 ± 0.003	0.048 ± 0.007
k_{cat} (s^{-1})	217.1 ± 10.1	162.1 ± 5.9	183.8 ± 8.6
k_{cat}/K_m (s^{-1} mM^{-1})	2390 ± 174	2600 ± 47	3820 ± 397
r-Pepsin (gly.)			
K_m (mM)	0.077 ± 0.002	0.067 ± 0.007	0.045 ± 0.006
k_{cat} (s^{-1})	181.4 ± 5.6	178.1 ± 3.2	124.0 ± 7.3
k_{cat}/K_M (s^{-1} mM^{-1})	2360 ± 141	2670 ± 311	2760 ± 237
c-Pepsin			
K_m (mM)	0.062 ± 0.010	0.054 ± 0.013	0.040 ± 0.001
k_{cat} (s^{-1})	115.8 ± 9.2	105.7 ± 10.4	105.6 ± 1.1
k_{cat}/K_M (s^{-1} mM^{-1})	1900 ± 177	2010 ± 268	2640 ± 46

[a] Kinetic constants were calculated based on a minimum of three measurements. No significant difference was found between r-pepsin and c-pepsin at the various pH values using the Student's t-test at the 0.05 level of significance.
Source: Ref. 14.

In conclusion, different protein expression systems can be successfully employed depending on the objectives of the research.

III. RATIONALE OF PROSEGMENT RESEARCH

Synthesis of proteolytic enzymes, including aspartic proteinases (APs), as inactive precursors is one way of regulating proteolytic activity in order to prevent undesirable degradation of proteins. These inactive precursors, known as zymogens, differ from the corresponding active enzymes by the presence of additional amino acids (referred to as the prosegment) at the N-terminal end of the protein. Conversion of the zymogen to its active form requires proteolytic removal of the prosegment hydrolysis of one or more peptide bonds either by the enzyme itself or by a second proteolytic enzyme [15]. Although it is known that the prosegment of APs is essential to maintaining the enzyme in its active form, little is known about the role of the prosegment in protein folding (both of the zymogen and activated enzyme) or the mechanism of activation [16].

Under physiological conditions, porcine pepsin is secreted in the stomach as a zymogen, pepsinogen. In pepsinogen, a 44-residue peptide, referred to as a prosegment, blocks the entrance to the active site and is held in place by electrostatic interactions with the pepsin molecule. During the conversion

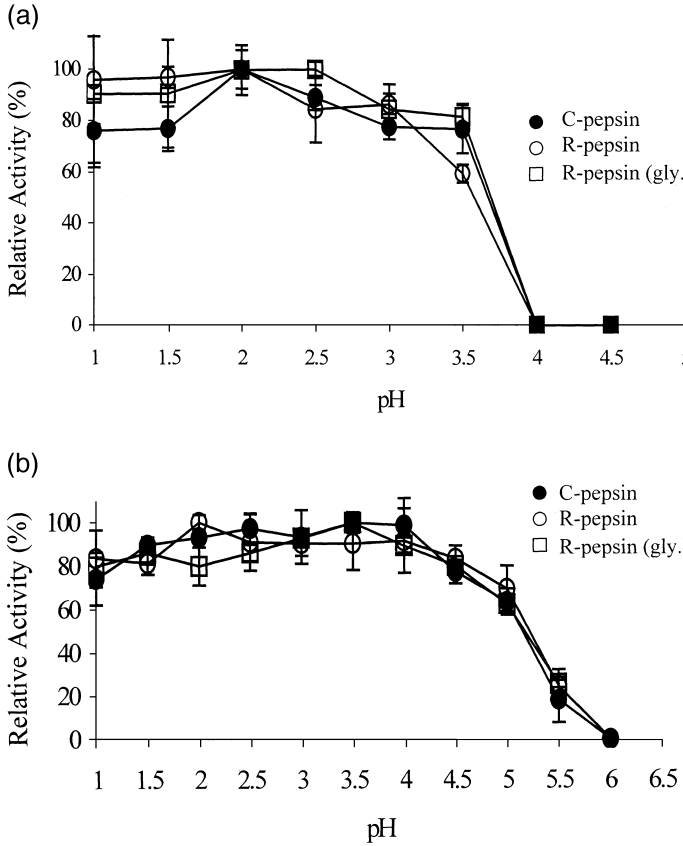


Figure 3 Effect of pH on activation of commercial and recombinant (nonglycosylated and glycosylated) pepsins (a). Effect of pH on activity of commercial and recombinant (nonglycosylated and glycosylated) pepsins (b). Hydrolysis of the synthetic substrate was measured at 300 nm over various pH values. The relative activity was expressed as a percentage of the highest activity over the pH range examined. Each data point represents the mean of three determinations. (From Ref. 14.)

from pepsinogen to pepsin, the prosegment is removed. The prosegment in pepsinogen covers the active site, although the actual cleavage site between enzyme body and the prosegment is not in the active site. The removal of the prosegment must occurs in conjunction with a conformational change in order to bring the digested position into the active site [17].

Possible rate-limiting conformational changes are the unmasking of the active site, the placement of the scissile bond into the active site, the movement

Table 3 First-Order Rate Constants (k_1) for the Activation of Wild-Type and K36pR Trx-PG at 14°C Determined with the Synthetic Peptide Substrate, KPAEFF(NO_2)AL

	First-order rate constant (min^{-1}) ± SD		
pH	1.1	2.0	3.0
Wild type	1.10 ± 0.10	0.519 ± 0.058	0.109 ± 0.011
K36pR	5.80 ± 0.51	1.25 ± 0.13	0.185 ± 0.025
K36pM	10.3 ± 1.2	6.56 ± 0.64	0.675 ± 0.012

Note: Standard deviations were calculated from three determinations.
Source: Ref. 20.

of the N-terminal 13 residues of pepsin from the active site to the first strand of the 6-stranded β-sheet, and the dissociation of the digested prosegment. In pepsinogen, Lys36p forms salt bridges with the catalytic aspartic acid residues. This salt bridges seem to be one of the key factors governing the stability of the prosegment [18]. Therefore, it is possible that replacing Lys36p with Glu, Arg, and Met will increase the rate of a conformational change required for activation by affecting the dissociation of the prosegment. Glick et al. [19] detected a rate-limiting conformational change that occurred at the same rate as the cleavage of the prosegment. This conformational change immediately preceded cleavage of the prosegment, limited the rate of prosegment cleavage, and was proposed to be the placement of the scissile peptide bond of the prosegment into the active site.

In light of the above, we undertook a study in which three mutants were constructed [K36pE (to see the effect of an opposite charge), K36pM (to see the effect of a hydrophobic charge), and K36pR (to see the effect of a similarly charged residue)] using the thioredoxin–pepsinogen fusion protein system (Trx-PG) [20]. K36pM and K36pE mutants were extremely unstable and

Table 4 Kinetic Parameters for Pepsins Obtained from the Activation of Wild-Type Trx-PG and K36pR, K36pM, and K36pE Trx-PG

	KPAEFF(NO_2)AL		LSF(NO_2)NleAL-OCH_3	
Substrate pepsin	K_m (mM)	k_0 (s^{-1})	K_m (mM)	k_0 (s^{-1})
Wild type	0.084 ± 0.019	88.8 ± 10.4	0.023 ± 0.004	49.1 ± 3.8
K36pR	0.050 ± 0.011	24.7 ± 2.4	0.023 ± 0.004	18.3 ± 1.4
K36pM	0.077 ± 0.010	30.6 ± 2.0	0.014 ± 0.002	19.6 ± 0.9

Note: Each value is the mean of three determinations.
Source: Ref. 20.

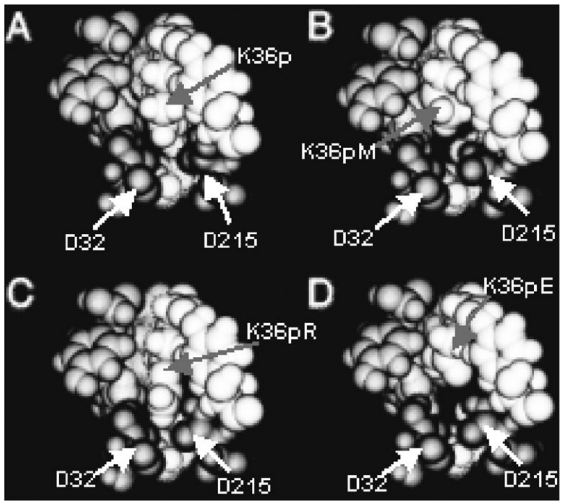


Figure 4 Space-filling diagrams of the active site of wild-type, K36pR, K36pM, and K36pE pepsinogen showing the packing of residue 36p and the catalytic aspartates. Wild type, (A); K36pR, (C); K36pM, (B); K36pE, (D); Asp 32 = D32, Asp215 = D215. (From Ref. 20.)

degraded rapidly, especially K36pE, which was inactivated during purification. This instability was confirmed by microcalorimetry where the denaturing temperatures for K36pM and K36pE were 6°C and 10°C lower than the wild type, respectively. As a function of the pH, K36pM and K36pR were activated over a broader range of pH as compared to the wild type. The mutant pepsinogens were activated faster than the wild type, with K36pM being activated ~10 times faster (Table 3). The activated pepsins from the various mutant pepsinogens showed lower kinetic efficiency than the wild type enzyme. Catalytic rate constants were reduced by half (Table 4). The results suggested that Lys36p is important for the correct folding of the active-center residues. The molecular modeling calculation suggested that the position of Asp 215 (one of the active-site residues) was substantially altered (Fig. 4). In conclusion, the above results would suggest that Lys36p was important not only for stability of the prosegment and pepsinogen but also for the correct alignment of the active-center residues.

IV. RATIONALE OF FLAP LOOP REGION

Loop structures have been identified in many enzymes where they have been implicated in protein function, stability, and possibly in protein folding.

Loops represent a relatively new category of secondary structures. They reverse the direction of a polypeptide chain and are usually situated at the protein surface. Unlike α-helices and β-sheets, loop structures have no regular patterns of dihedral angles and hydrogen bonds [21,22]. Loop residues that are involved in catalysis may be more highly conserved than amino acids in surface loops whose role is purely structural. This sequence conservation can involve amino acids that interact directly with a substrate as well as amino acids that help stabilize various conformations that the loop may adopt during catalysis [23]. The hydrogen bonds to main-chain polar atoms play an important role in maintaining medium-size loop structures. They can be formed from the loop atoms to main-chain or side-chain atoms or ligands [24]. It was observed in lactate dehydrogenase that open-loop and closed-loop conformations were facilitated by the same pattern of hydrogen bonds made among loop atoms [25]. In contrast, hydrogen bonds in triosephosphate isomerase reorganized during catalysis [26].

In aspartic proteinases, a loop commonly known as a flap is found to extend over the active site and has been implicated in the catalytic process [27,28]. Initial X-ray diffraction studies have shown that there is a very flexible β-hairpin flap located at the entrance of the active site (residues 71–82). X-ray crystallography data of various APs have shown that the flap contributes to the S_2, S_1 and S_2' binding pockets of the enzymes [29]. The binding of pepstatin, an inhibitor of aspartic proteinases, has resulted in rigidification and stabilization of the flap and the substrate (rhizopepsin [30], penicillopepsin [31], pepsin [32]). The flap thus ensures an environment secluded from the solvent such that transient substrate intermediates are encapsulated by the flap and stabilized [33]. Three flap residues contribute directly to substrate subsite specificity. Thr77 contributes to S_1 and S_2 subsites, Gly76 to the S_2 subsite, and Tyr75 to S_1 subsite [18].

Of the above three residues, Tyr75 has been studied the most. Tyr75 is conserved in all eukaryotic APs and is located near the tip of the flap [17,28]. Its phenolic ring forms one of the walls of the S_1 substrate-binding pocket. Despite the large displacement of the flap, the position of Tyr75 is unusually stable in both the liganded and free forms of APs with the exception of chymosin [34]. Site-directed mutagenesis has been used to convert Tyr75 of chymosin to Phe [35] as well as to convert this residue in *Mucor pusillus* pepsin to Phe and Asn [36].

Gly76 has been implicated in the catalytic process and it is thought to participate in the hydrogen-bond network aligning the substrate. Glycine residues are known to contribute to conformational flexibility of polypeptide chains and have been found to contribute to the flexibility of some loops associated with catalysis. A comparison of porcine pepsin in zymogen, mature and inhibited forms, revealed that a loop (a flap), consisting of residues 71–80,

Table 5 Comparison of Milk Clot and Proteolytic Activities

Pepsin	Milk clot (MC)	Proteolysis (Pa)	MC/Pa
Wild type	1.56 ± 0.03	0.54 ± 0.00	2.9
Y75F	1.54 ± 0.03	0.89 ± 0.01	1.7
Y75N	0.86 ± 0.07	0.15 ± 0.01	5.7

MC = amount of the protein which gives 0.4 absorbance change in one min.
Pa = amount of the protein which gives one absorbance of soluble peptide in one min.
Source: Ref. 41.

located near the active site changed its position upon substrate binding. The loop residue, glycine 76, has been implicated in the catalytic process and thought to participate in a hydrogen-bond network aligning the substrate [37–39].

Thr77 is replaced by an aspartate in fungal enzymes and is thought to be responsible for greater substrate specificity as compared to the mammalian enzymes [40].

Although investigated by other researchers in other aspartic proteinases, the exact roles of Tyr75, Gly76, and Thr77 in pepsin had not been defined. Therefore, we undertook site-directed mutagenesis studies to investigate their roles in structure–function relationships of pepsin(ogen).

In a study to examine the role of Tyr75, this residue was mutated to Phe (which lacks a *p*-hydroxyl group) and to Asn (which resembles the bottom phenolic ring of Tyr) [41]. The mutations resulted in localized alterations to the active site region and was reflected in a change of the catalytic rate (Tables 5 and 6). The pH dependency and circular dichroism (CD) analyses indicated (i.e., no change among samples), however, that the effect was not global. Molecular modeling suggested that the mutations allowed the loop to assume another position, thereby changing the catalytic efficiency (Fig. 5).

The role of glycine 76 was investigated by mutating Gly to Ala, Val, and Ser [42]. These amino acids differ in their van der Waals volumes, accessible

Table 6 Kinetic Constants for Pepsins

Pepsin	K_m (mM)	k_0 (s^{-1})
Wild type	0.034 ± 0.005	65.4 ± 3.1
Y75F	0.025 ± 0.003	82.6 ± 3.3
Y75N	0.036 ± 0.006	12.2 ± 0.7

Source: Ref. 41.

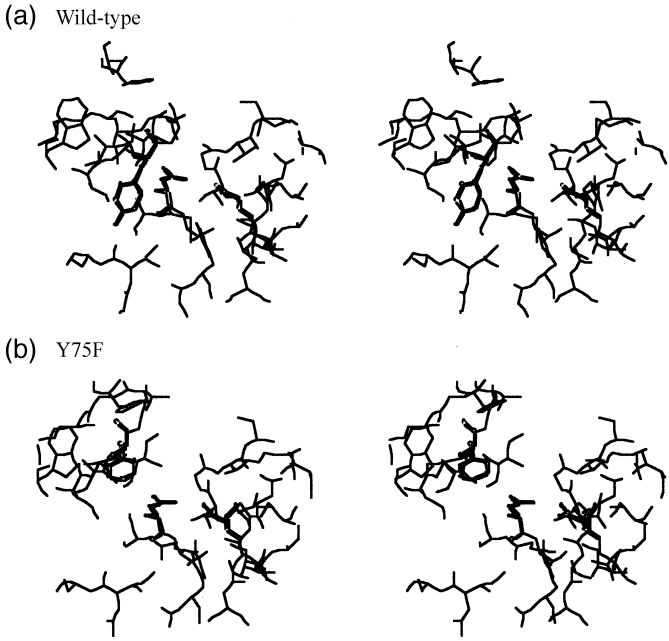


Figure 5 Comparison of the active-site cleft conformations. (a) Wild-type pepsin at the maximum opening of the loop; (b) Y75F pepsin at the maximum opening of the loop. The loop locates on the right side of each figure. Both catalytic aspartate residues and residue 75 are shown in ball-and-stick model. The entrance of the cleft opens wider in Y75F than WT (4.1 Å versus 2.5 Å). (From Ref. 41.)

Table 7 Kinetic Parameters for the Wild-Type Pepsin and G76 Mutants

Substrate	Enzyme	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1} s^{-1})
KPAEFF(NO_2)AL	Wild type	0.043 ± 0.005	180.5 ± 10.3	4198 ± 544
	G76A	0.098 ± 0.008	47.0 ± 9.9	395 ± 96
	G76V	0.071 ± 0.011	11.1 ± 1.1	160 ± 30
	G76S	0.085 ± 0.005	2.5 ± 0.6	29 ± 7
LSF(NO_2)NleAL-OCH_3	Wild type	0.011 ± 0.002	88.7 ± 4.6	8064 ± 1525
	G76A	0.040 ± 0.009	43.4 ± 3.6	1085 ± 260
	G76V	0.044 ± 0.005	1.2 ± 0.1	28 ± 3
	G76S	0.041 ± 0.008	2.9 ± 0.3	70 ± 16

Note: Kinetic constants were calculated based on minimum two measurements for each enzyme.
Source: Ref. 42.

Table 8 First-Order Rate Constants for Activation of Thioredoxin–Pepsinogen Fusion Proteins

	k (min^{-1})		
Enzyme	pH 1.1	pH 2.0	pH 3.0
Wild type	4.449 ± 0.262	2.819 ± 0.152	1.043 ± 0.089
G76A	0.311 ± 0.076	0.332 ± 0.014	0.104 ± 0.029
G76V	0.132 ± 0.001	0.152 ± 0.013	0.04 ± 0.001
G76S	0.130 ± 0.014	0.166 ± 0.014	0.059 ± 0.003

Note: Rate constants were calculated based on three measurements for each enzyme and pH.
Source: Ref. 42.

surface areas, polarities, and allowable energy levels on Ramachandran plots for individual amino acids [43]. In addition, the serine mutant introduced a hydrogen-bonding potential at position 76, similar to that observed in human renin. All of the mutants, regardless of amino acid size and polarity, had a lower catalytic efficiency (Table 7) and activated more slowly than the wild-type enzyme (Table 8). The slower activation process was associated indirectly with altered proteolytic activity. Consequently, it was proposed that a proteolytic cleavage represents a limiting step of the activation process. Lower catalytic efficiency of the mutants was explained as a decrease in the flap flexibility and, therefore, a different pattern of hydrogen bonds responsible for substrate alignment and flap conformation. The results demonstrated that flap flexibility is essential to efficient catalytic and activation processes [40].

Table 9 First-Order Rate Constants for Activation of Thioredoxin–Pepsinogen Fusion Protein T77 Mutants

	k (ms^{-1})		
Enzyme	pH 1.1	pH 2.0	pH 3.0
Wild type	36.5 ± 6.7	15.6 ± 1.1	7.8 ± 0.5
T77S	32.0 ± 4.6	15.2 ± 1.7	4.5 ± 0.4
T77V	12.3 ± 1.1	9.8 ± 1.1	2.9 ± 0.8
T77G	11.8 ± 1.9	8.4 ± 0.2	1.6 ± 0.1

Note: Rate constants were calculated based on three measurements for each enzyme and pH.
Source: Ref. 44.

Table 10 Kinetic Parameters for the Wild-Type and T77 Pepsin Mutants

Substrate	Enzyme	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1} s^{-1})
KPAEFF(NO_2)AL	Wild-type	0.043 ± 0.005	180.5 ± 10.3	4198 ± 548
	T77S	0.052 ± 0.008	197.7 ± 11.0	3802 ± 622
	T77V	0.096 ± 0.010	105.9 ± 5.5	1103 ± 128
	T77G	0.074 ± 0.014	18.7 ± 1.5	253 ± 52
LSF(NO_2)NleAL-OCH_3	Wild-type	0.011 ± 0.002	88.7 ± 4.6	8064 ± 1525
	T77S	0.018 ± 0.004	102.8 ± 5.2	5711 ± 1302
	T77V	0.022 ± 0.003	8.8 ± 0.4	400 ± 57
	T77G	0.037 ± 0.009	2.7 ± 0.8	73 ± 28

Note: Kinetic constants were calculated based on minimum two measurements for each enzyme.
Source: Ref. 44.

Finally, with regard to Thr77, three mutants, T77S, T77V, and T77G, were constructed, differing in polarity and hydrogen-bonding potential [44]. T77S activated at the same rate and had similar catalytic parameters as the wild-type pepsin. The activation rates of T77V and T77G were slower and the catalytic efficiencies lower than the wild type (Tables 9 and 10). The contribution of the side chain to zymogen activation was associated with the proteolytic cleavage of the prosegment. It was postulated that the hydroxyl group at position 77 provided an essential hydrogen bond that contributed to proper substrate alignment and indirectly to catalytically favorable geometry of the transition state [44].

V. STABILITY OF THE PROTEIN

One of the goals of genetic engineering is to understand factors responsible for the destabilization of protein structure (i.e., denaturation) and, ultimately, the loss of function as a result of changes in environmental conditions (e.g., heat, pH) [45]. Therefore, a challenging goal of protein engineering is the design of proteins with significantly enhanced chemical and thermal stabilities. Although several strategies have been attempted, their pursuit has often yielded proteins that are less stable than the wild-type (WT) progenitors for reasons which are not fully understood [46]. Glycosylation in proteins, through posttranslational modification, has been proposed to have several functions: resisting proteases, increasing thermal stability, preventing aggregation of proteins, and maintaining proper conformation for activity [47]. Yet, the exact nature of the contribution of attached carbohydrates to protein structure–function relationships has not been clearly elucidated, espe-

cially for aspartic proteinases. Of the aspartic proteinases studied to date, only *Mucor miehei* proteinase is known to be glycosylated and is known for its high thermal stability [48]. Pragmatically, if the stability of proteinases to various environmental conditions could be expanded, then enzymes could be more readily used as processing aids given that many food systems are exposed to a wide range of processing conditions (pH and temperatures) [49]. From a fundamental/molecular point of view, understanding those regions/amino acids responsible for stability is paramount in developing strategies to alter protein functionality.

All proteins are synthesized in a neutral pH environment; thus, their natural conformation state and functionality exists in this environment. However, most aspartic proteinases are stable under acidic conditions and become irreversibly denatured under neutral pH conditions [50,51]. In the case of pepsin, its zymogen, pepsinogen, is stable under neutral pH conditions. Because the majority of the pepsin molecule is identical to pepsinogen, minor differences may be responsible for pepsin's instability at neutral pHs.

Why is pepsin so instable despite being similar in conformation to pepsinogen? The most obvious differences between pepsin and pepsinogen are found in the prosegment portion. Not only does the prosegment cover the active-site cleft, but it contains a large number of positively charged residues [i.e., 13 (9 Lys, 2 Arg, 2 His) out of the 44 residues in the prosegment of pepsinogen are positively charged] [52,53]. Pepsin, on the other hand, contains only four positive residues (two Arg, one Lys, one His). Therefore, the number and distribution of these positively charged residues contribute to the difference in pH stability between pepsin and pepsinogen. In contrast, unlike most aspartic proteinases, chicken pepsin is relatively stable at neutral pHs [50]. A homology search between porcine and chicken pepsin shows a 74% similarity. Major differences between chicken and porcine pepsins are in charge distribution and the N-terminal amino acid sequences. In chicken pepsin, there are 21 Asp, 13 Glu, 9 Lys, 5 Arg, and 4 His charged residues, whereas in porcine pepsin, 28, 13, 1, 2 and 1 are charged, respectively. For both porcine and chicken pepsin, most of these charged residues are distributed on the surface; however, the distribution of the positively charged residues in chicken pepsin is more favorable to neutral conditions.

Another difference between pepsin and pepsinogen is the location of the N-terminal fragment. In the zymogenic form and during activation, the N-terminal fragment is in the active-site cleft. After activation, this N-terminal fragment is placed in the β-sheet at the bottom of the protein. This relocation results in about a 40-Å movement [52]. Because this portion relocates from one side of the protein to other side, it is suggested that this fragment could be readily moved from its position and could be important for the stabilization

of pepsin under neutral pH conditions. Lin et al. [54] reported that denaturation of pepsin begins with the denaturation of the N-terminal domain and suggested that the interaction between the N-fragment and the enzyme body is important for stability. A recent crystallographic study of cathepsin D, another neutral pH stable aspartic proteinase, showed that the N-terminal fragment of this enzyme is relocated to its active-site cleft at pH 7.5, resembling its zymogenic form [55].

From the above observations, several key factors in the instability of pepsin are suggested. In this study, the following were undertaken in order to investigate and possibly improve the stability of pepsin at neutral pH's: (1) the alteration of the charge distribution on the surface of the pepsin; (2) the deletion of the extra amino acid sequence observed in porcine pepsin C domain; (3) the addition of hydrogen-bond stabilizers; and (4) the stabilization of the N-terminal portion of pepsin through mutations and the introduction of a disulfide bond [56].

VI. CHARGE DISTRIBUTION

Replacements of the positive residues were chosen in the region where negatively charged residues were highly concentrated. Mutations were selected in order to resemble the sequence of chicken pepsin. In the resultant chicken-like mutant, N + C, seven positive residues were added and three negative residues were removed. Kinetic measurements of this mutant were taken at pH 2.1 and 3.95 and shown to have comparable kinetic constants to the wild type (Table 11). The mutant inactivated at about half the rate of the wild type. The inactivation rates of the mutants which had mutations on only

Table 11 Kinetic Constants of Mutant and Wild-Type Pepsin

	ss1 (KPAEFF(NO_2)AL)		ss2 (LSF(NO_2)NleAL methyl ester)	
Enzyme	K_m (mM)	k_0 (s^{-1})	K_m (mM)	k_0 (s^{-1})
Wild type	0.034 ± 0.005	65.4 ± 3.1	0.019 ± 0.006	190.5 ± 18.9
N + C	0.057 ± 0.005	77.5 ± 2.9	0.031 ± 0.013	50.7 ± 9.4
Δ240-246	0.079 ± 0.006	60.9 ± 2.6	0.015 ± 0.002	25.7 ± 1.0
N-frag	0.083 ± 0.015	68.9 ± 6.6	0.028 ± 0.007	80.7 ± 10.2
N-frag (A)	0.075 ± 0.008	186.0 ± 9.8	0.028 ± 0.007	122.0 ± 15.6
N-frag (B)	0.065 ± 0.009	187.1 ± 12.8	0.041 ± 0.005	315.3 ± 16.8
G2C/L167C	0.080 ± 0.007	41.6 ± 2.0	0.018 ± 0.001	13.0 ± 0.22

Note: Values represents the mean of three determinations ± standard deviations.
Source: Ref. 56.

Table 12 Rate Constants of Inactivation of Mutants at pH 7.0

Enzyme	k_d (min^{-1})
N-frag	0.0169 ± 0.0014
N-frag (glycerol, sucrose)	0.00229 ± 0.00031
N-frag (pH 7.5)	0.268 ± 0.008
G2S/D3Y (N-frag(A))	0.0680 ± 0.0038
L10M/T12A/E13S (N-frag(B))	0.0646 ± 0.0015
G2C/L167C	0.0536 ± 0.0051
N + C	0.0421 ± 0.0071
N-DOM	0.0905 ± 0.0173
C-DOM	0.0852 ± 0.0237
Wild type	0.0991 ± 0.0075
Wild type (glycerol)	0.0365 ± 0.0026
Wild type (sucrose)	0.0203 ± 0.0009
Wild type (glycerol, sucrose)	0.00743 ± 0.00216
Wild type (pH 7.5)	ND (too rapid)

Note: ND: Inactivation was too rapid (completed within 5 min) under the experimental conditions to determine the inactivation rate. Values represents the mean of three determinations ± standard deviations.
Source: Ref. 56.

the N- or C-terminal domains (N-DOM and C-DOM, respectively) were comparable to that of the wild type. Therefore, mutations on the surface had little effect on the kinetic parameters (Table 12).

Although the N + C mutant had similar activity as the wild type, the stability at neutral pH changed (Table 12, Fig. 6). Because Lin et al. [54] had shown that the initial denaturation occurred in the N-terminal domain, the mutations on the N-terminal domain were expected to have a dominant effect. The negatively charged residues on the N-terminal domain themselves, however, only had a minor effect on the stability. Thus, the results of the above three mutants demonstrated that the distribution of the negatively charged residues on the surface of the porcine pepsin helped to stabilize the enzyme; however, the degree of stabilization was not substantial [53].

VII. N-TERMINUS

The position of the N-terminal fragment in pepsin changes from the active-site cleft to the opposite side of the molecule during activation [52]. This implies that the N-terminal portion is relatively easily displaced from its

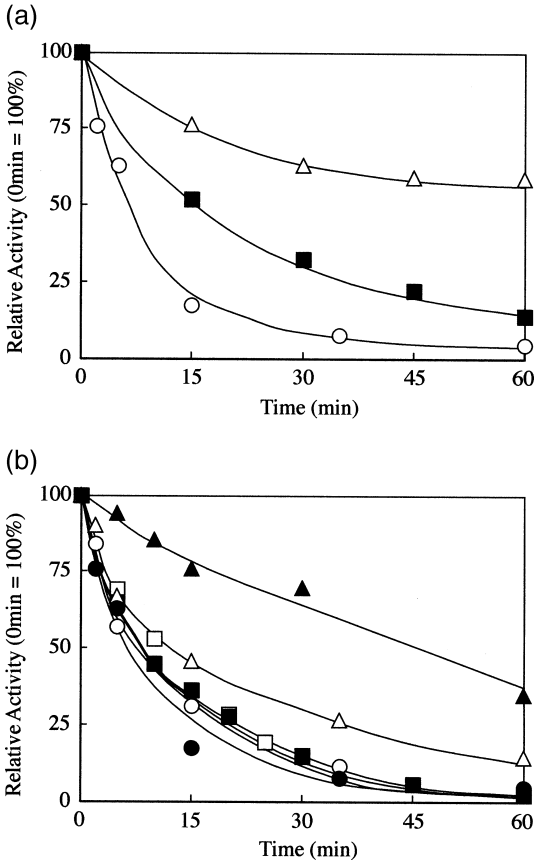


Figure 6 Inactivation of wild-type and mutant pepsins. Pepsin was inactivated in the conditions described in the text and residual activities were plotted. (a) Open circle, wild type; closed square, wild type with glycerol; open triangle, wild type with glycerol and sucrose. (b) Closed circle, wild type; open triangle, N + C; closed triangle, N-fragment; open circle, Δ240–246; open square, N-frag (a); closed square, N-frag (b). Each data point represents the mean of three determinations. (From Ref. 56.)

original position. If this N-terminal portion moves away from its original position, a β-sheet loses one of its strands, raising the possibility of destabilization of the sheet and/or interaction with another hydrogen-bond source. Therefore, it is suggested that stabilizing this fragment in its original position, which was a strand of a β-sheet on the bottom of pepsin molecule, will prevent denaturation in the event of neutralization. In order to examine this possi-

bility, mutations were introduced into the N-terminal portion to keep it in the fixed position at the bottom of the pepsin molecule.

The comparison of the amino acid sequence of the pepsins from porcine and chicken (chicken pepsin is relatively stable at neutral conditions) revealed major differences in the N-terminal portion. We chose five amino acid residues to be replaced. The mutation of the N-terminal portion was done with two sets of mutations, G2S/D3Y (N-frag(A)) and L10M/T12A/E13S (N-frag (B)). These two mutants and a third mutant which combined both mutations (N-frag) exhibited kinetic constants similar to the wild type and had comparable catalytic activities for both synthetic substrates (Table 11).

Stability tests of the N-frag mutant showed that it was 5.8 times more stable than the wild type (Table 12, Fig. 6). Although the wild-type was inactivated in 60 min, this mutant retained 30% of its activity after 60 min. Even after 4 h, 18% of the original activity was observed with this mutant. When the pH was raised to 7.5 (Fig. 6), the wild type showed no activity after 5 min. Even the addition of glycerol and sucrose, which was shown to stabilize the enzyme, had no effect. When the N-frag mutant was exposed to pH 7.5, it retained 5% activity after 15 min. The rate constant of inactivation for N-frag at pH 7.5 was 0.268 min^{-1}. Moreover, in the presence of both glycerol and sucrose at pH 7.5, the enzyme retained 25% activity after 30 min.

Because the N-frag mutant had five amino acid replacements, some of the mutations could be more critical than others. The G2S/D3Y and L10M/T12A/E13S mutants, however, showed less stability than the N-frag mutant (Fig. 7). Both mutants showed an inactivation rate which was about 1.5 times slower than the wild-type while the N-frag mutant was 5.8 times slower. Also, at pH 7.5, the activities of both N-frag(A) and N-frag(B) mutants were quenched slower than the wild type, but faster than the N-frag mutant. After 5 min, N-frag(A), N-frag(B), N-frag, and wild type had 0%, 3%, 27%, and 0% of the original activity in the absence of glycerol and sucrose whereas 45%, 46%, 0% and 72% of the original activities remained in the presence of glycerol and sucrose, respectively.

Molecular model minimization showed that these mutations did not contribute to the internal interactions (Fig. 8). The only major difference between the wild-type and N-frag was the addition of a hydrogen bond between Ser2 γ-oxygen and L167 nitrogen in the G2S mutation. Kinetics of the N-frag(A) mutant, however, showed this addition was insufficient to stabilize the protein entirely. It was, thus, concluded that each of the five replacements by themselves were not critical, but synergistically helped to stabilize the enzyme.

Because each of the above mutations in itself was not critical to stabilization, the mechanism of stability was still in question. The authors suggest two possibilities on how the mutations stabilized pepsin: (1) the release of the N-

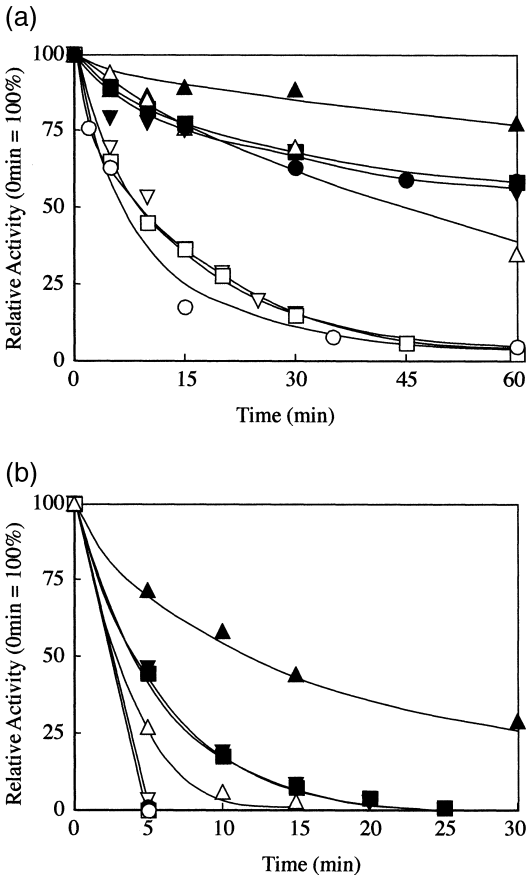


Figure 7 Inactivation of the N-terminal fragment mutant. Inactivation test of N-fragment, N-frag(A) and N-frag(B) mutants showed that individual mutations in the N-fragment mutant were not critical for the stabilization of pepsin. Both N-frag(A) and N-frag(B) mutants showed slight stabilization effects (B) and were similar in stability (A) to the wild-type while the N-frag mutant showed drastic stabilization. (a) The inactivation reactions were carried out at pH 7.0. (b) The inactivation reactions were carried out at pH 7.5. Symbols: wild type (open circle); N-fragment (upright open triangle); wild-type with glycerol and sucrose (closed circle); N-frag(A) with glycerol and sucrose (inverted closed triangle); N-frag(B) with glycerol and sucrose (closed square); N-fragment with glycerol and sucrose (upright closed triangle). Each data point represents the mean of three determinations. (From Ref. 56.)

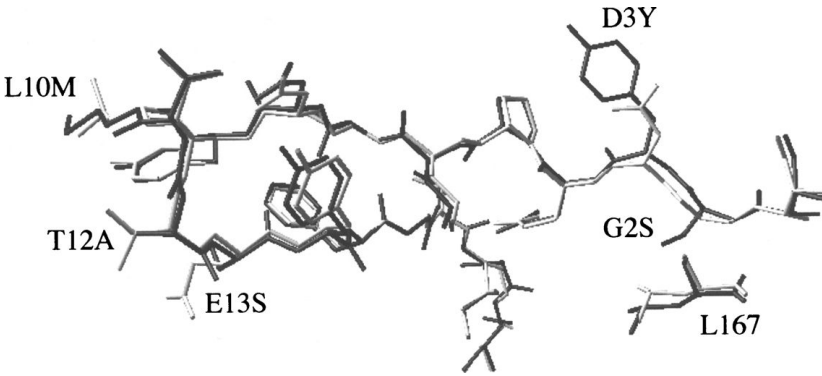


Figure 8 Molecular model comparison of wild-type and N-terminal fragment mutant pepsins. Superposition of the N-terminal fragment of wild-type (gray) and the N-frag mutant (black) models. There were no major differences. The root-mean-square deviation between backbone atoms of wild-type and N-frag was 0.28 Å. The numbers show the mutation sites in N-frag. (From Ref. 56.)

terminus portion is prevented with these mutations or (2) these mutation sites are responsible for the stability of the released N-terminal portion. Crystal structure of inactive cathepsin D at pH7.5 showed that the N-terminal portion is relocated into the active-site cleft and is stabilized by an interaction to the catalytic site [55]. From this crystal structure data, the authors suspect that the second posibility is more likely.

In a follow-up to the above study, research was undertaken to engineer a pepsin form that displayed neutral pH stability utilizing the structural motif Lys-X-Tyr on the N-terminus capable of the stabilizing mechanism observed in cathepsin D [57]. Cathepsin D exhibits unusually high neutral pH stability compared to most aspartic proteinases and a crystallographic study revealed the structural mode of stabilization [55]. The pH 7.5 structure of cathepsin D revealed that the N-terminus had relocated to the active site by a 180° rotation, where it was bound via a salt bridge between Lys8 and the catalytic Asp residues [55]. It was also observed that Ala13 functioned as a hinge point for rotation of the N-terminus into the active site. It was suggested that the Lys-X-Tyr motif of the cathepsin D N-terminus played the role of prosegment residues Lys36p and Tyr9 (pepsin numbering) in the aspartic proteinase zymogens [55]. Three mutations were designed upon comparison of the N-terminal sequences of Cathepsin D and pepsin. Lys was incorporated at position 7 (equivalent to position 8 of cathepsin D) in place of Glu. Lys at position 7 was to provide for the formation of a salt bridge with the catalytic Asp residues, as observed in cathepsin D. Thr12, the positional equivalent to

Ala13 in cathepsin D, was replaced with Ala to provide a hinge point for rotation of the N-terminus into the active site. Finally, Glu13, equivalent to Gln14 in cathepsin D, was changed to Gln to avoid charge–charge repulsion with other carboxyl groups ionized at neutral pH, potentially inhibiting the 180° rotation about the hinge point. The mutations were located as follows: E7K is part of the reverse turn, and Glu7 normally interacts via a hydrogen-bond pair with Phe15 of the neighboring sheet. T12A is part of the loop within the reverse turn that connects the sixth strand of the interdomain β-sheet with the first strand of a neighboring sheet [58]. The Thr12 side chain shares a H-bond with the main-chain carbonyl group of Leu10 [17]. E13Q is the last position of the same reverse turn and Glu13 shares a β-sheet interaction with Tyr9, the first residue of the reverse turn [18,58].

At pH 7.0, E7K/T12A/E13Q pepsin was inactivated more slowly compared to WT, whereas the mutants E7K and T12A/E13Q were inactivated similar to WT (Fig. 9). Far-UV ultraviolet circular dichroism revealed that changes in secondary structure accompanied the inactivation process and that the structural changes occurred at approximately the same rate as inactivation (Fig. 10). All of the inactivated pepsin forms showed retention of

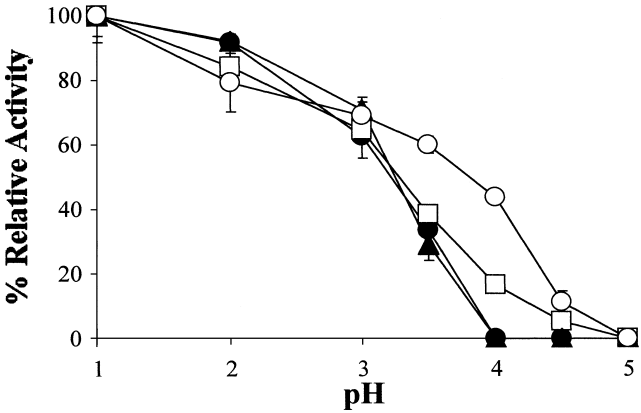


Figure 9 The pH dependence of activation of the zymogen fusion protein samples as indicated by percent relative activity after incubation at the various pH values shown. Closed circles, wild-type; squares, E7K; triangles, T12A/E13Q; open circles, E7K/T12A/E13Q. Activation reactions contained 250 n*M* protein and 100 m*M* sodium citrate buffer and were incubated at 25°C. Activity was assessed after exactly 5 min incubation by the proteolytic assay described in [57] for the synthetic substrate KPAEFF(NO2)AL using a 12.5 n*M* enzyme and 100 μ*M* substrate. Activity was expressed in terms of percent relative activity with respect to the highest rate obtained for each pepsin form, respectively. (From Ref. 57.)

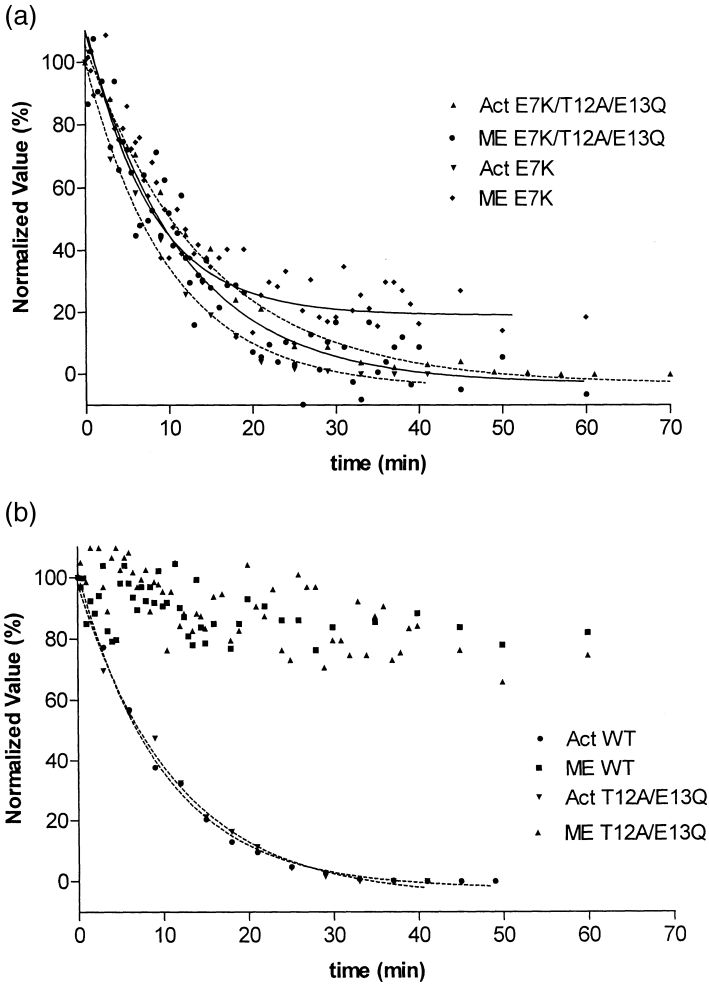


Figure 10 Comparison of the normalized changes in molar ellipticity at 222 nm (ME; solid lines) and loss of proteolytic activity (Act; broken lines). Incubation was at room temperature in 100 m*M* sodium phosphate buffer, pH 7.0, containing 20% glycerol. Nonlinear regression for an exponential decay and unpaired *t*-testing of the rate constants were performed. (a) The exponential rate constants for molar ellipticity change and activity loss were not significantly different for E7K ($p = 0.36$) and E7K/T12A/E13Q ($p = 0.15$), respectively. (b) The lack of a relationship between molar ellipticity change and activity loss is shown for WT and T12A/E13Q. The molar ellipticity at 222 nm for WT and T12A/E13Q did not vary with time linearly or nonlinearly in a significant fashion. (From Ref. 57.)

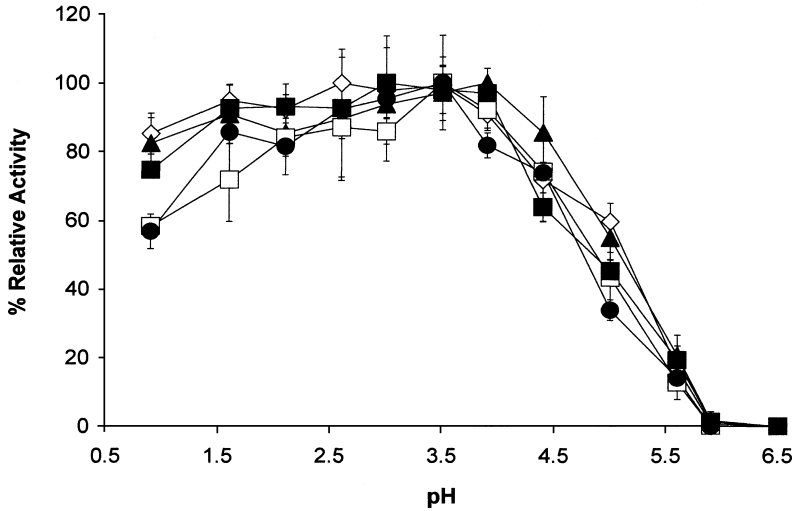


Figure 11 The pH dependence of activity profile for the various pepsin forms; WT (diamonds), E7K (triangles), T12A/E13Q (open squares), E7K/T12A/E13Q (circles), and commercially obtained pepsin (closed squares). Activity was assayed by the rate of cleavage of the chromophoric substrate KPAEFF(NO2)AL, 100 μM, 95 mM sodium citrate buffer, at room temperature, and 6–9 nM enzyme. (From Ref. 57.)

substantial secondary structure, more than previously determined for pepsin denatured at pH 7.2 and 8.0, suggesting the presence of a structural intermediate at pH 7.0. The coupled mutations at positions 12 and 13 impacted the pH dependence of activity at pH 0.9 (Fig. 11), altered specificity as determined by insulin B-chain cleavage profiles, lowered affinity for a synthetic substrate, and lowered the turnover number (Table 13). The introduction of Lys at

Table 13 Summary of the Calculated Kinetic Parameters Obtained by Nonlinear Regression Using the Method of Least Squares for Best Fit to the Michaelis–Menten Model

	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\mu M^{-1}s^{-1}$)
WT	70 ± 10	98 ± 7	1.4
E7K	100 ± 20	65 ± 6	0.65
T12A/E13Q	190 ± 30	42 ± 4	0.22
E7K/T12A/E13Q	190 ± 28	76 ± 7	0.39

Note: All values were obtained by averaging determinations done in triplicate.
Source: Ref. 57.

position 7 apparently destabilized the interaction between prosegment and enzyme body as evidenced by activation at higher pH ($\geq$4.0) compared to WT, altered specificity for insulin B-chain, but showed no change for pH-dependence of activity, nor a statistically significant change in affinity for the synthetic substrate.

VIII. DISULFIDE LINKAGE

If the mobility of the N-terminal portion initiates denaturation, then it would be logical to think that fixing this portion to the enzyme body would prevent denaturation. To fix the N-terminus portion to the enzyme body, a potential disulfide bond was introduced. This mutant, G2C/L167C, had a cysteine residue at the second residue of the N-terminal portion and another cysteine on the opposite side of the enzyme body (Fig. 12). The kinetic studies of this mutant showed a lower but substantial amount of activity as compared to the wild type (Table 11).

The rate constant of the inactivation of G2C/L167C at pH7.0 was about half of the wild type but was not as slow as the N-frag. However, inactivation slowed down after 30 min and had lower but noticeable activity (3.2%) over 24 h, whereas the N-frag, N-frag(A), and N-frag(B) mutants were completely inactive after 24 h. These results would imply that the formation of the

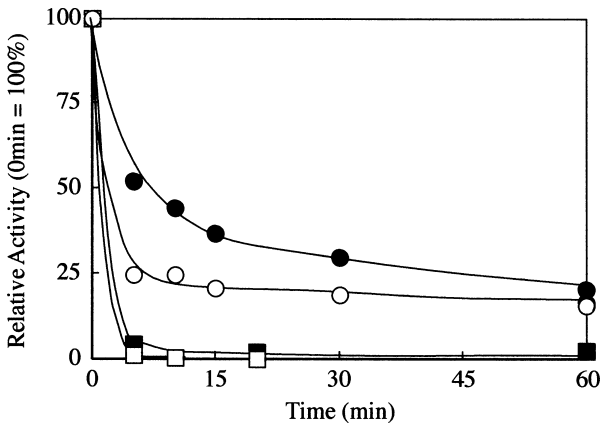


Figure 12 Inactivation of G2C/L167C mutant pepsin. Inactivation of G2C/L167C mutant and the effect of the oxidizing reagent are shown. At pH 7.5, the oxidization did not stabilize the wild type [without oxidization (open square) and with oxidization (closed square)]. Each data point represents the mean of three determinations. (From Ref. 56.)

disulfide bond prevented denaturation by preventing the movement of the N-terminal portion (Table 12). However, the slower disulfide bond formation seemed to compete with the faster inactivation because substantial activity was lost before reaching the plateau. To increase the rate of disulfide bond formation, oxidizing reagents were used (i.e., $FeCl_3$, $K_3Fe(CN)_6$, *o*-iodothobenzoate, dithionitrobenzoate, and 3,3′ -dithiopyridine). These reagents had comparable effects. After 24 h of oxidization with 2 m*M* $FeCl_3$, the oxidized G2C/L167C was tested for stability at pH 7.5 (Fig. 12). G2C/L167C, without oxidizers, showed slower inactivation than wild type; however, most of the activity was lost after 480 min. In the presence of $FeCl_3$, inactivation slowed down and reached a plateau at 20% of its initial activity. Further inactivation studies showed that activities were retained: 11% at 24 h and 5% at 74 h.

These results indicated that the disulfide bond formation kept the N-terminal fragment close to its native position, thereby stabilizing the enzyme. Formation of disulfide bonds was faster at pH 7.5 than at pH 7.0, thus, greater stabilizing effect were observed at the higher pH.

These results revealed pepsin's instability at neutral pHs are due to relocation of the prosegment. In fact, an aspartic protease, human cathepsin D, has a stabilization mechanism of the prosegment in neutral pHs. This enzyme does not prevent the dissociation of the amino terminal fragment, but it provides the accomodation of the dissociated amino fragment at another position under neutral conditions.

IX. POSTTRANSLATIONAL MODIFICATION

The role of ligands, especially carbohydrates, in protein structure–function studies has been widely recognized, although the specific role of glycosylation for a given protein has not been clearly elucidated.

Glycosylation, a posttranslational process where carbohydrate residues are covalently attached as the protein is being processed, has been found to facilitate proper folding, prevent protein aggregation, protect against proteolytic attack, and act as surface cell receptors and targeting signals [59–62]; however, specific roles appear to be protein dependent. For some glycoproteins, N-linked oligosaccharides are necessary for overall stability [63,64], whereas for others, the presence of the carbohydrate chains is only required during the folding process [65]. Some proteins fold more efficiently without carbohydrates, whereas partial removal of the oligosaccharide chains is essential for others [66]. For glycoproteins containing multiple N-linked motifs, some sites have more impact on the stability and function than others [67–69]. On the other hand, removal of the N-linked sites has occasionally had little effect on structure and function [70]. For example, staphylokinase, an enzyme that contains the N-linked motif but is not found in a naturally glycosylated state, was expressed from a eukaryotic system in a glycosylated

form, and in this case, glycosylation was shown to have no effect on the structure, stability, or activity [71].

Much of the research on the role of glycosylation has been devoted to naturally glycosylated proteins. Using site-directed mutagenesis and various expression systems, more systematic analysis of individual N-linked carbohydrate chains on glycoproteins has been achieved. It has allowed researchers to study multiple N-linked sites to determine the impact that each oligosaccharide has on the structure and function of the glycoprotein [67–69] Conversely, little research has been conducted on the effect of glycosylation on nonglycosylated proteins. Kato and co-workers [72,73] engineered N-linked sites into lysozyme with an intent to improve its emulsifying properties.

Given the above and the fact that neither pepsinogen nor pepsin are glycosylated, the effects of N-linked glycosylation on the structure–function relationship of porcine pepsin(ogen) was undertaken [74]. The asparagine-X-serine/threonine motif was engineered into four sites (two on the N-terminal domain, positions 77 and 110, and two on the C-terminal domain, positions 244 and 288) and the various glycosylated pepsins were analyzed for changes in structure (secondary and tertiary), enzyme activity, substrate specificity, and stability (pH, temperature). A study was undertaken. All four N-linked recombinants exhibited similar secondary and tertiary structure to nonglycosylated pepsin (i.e., wild type). Similar K_m values were observed, but catalytic efficiencies were approximately one-third for all mutants compared to the wild type (Table 14). Despite the difference in catalytic activity, sub-

Table 14 Kinetic Analysis of Glycosylated and Nonglycosylated Pepsins at pH 2.0

Enzyme	K_m(mM)	k_{cat} (s^{-1})	k_{cat}/K_m(s^{-1}mM^{-1})
Wild type	0.062 ± 0.003	162.1 ± 5.9	2600 ± 47
Mutant 77	0.052 ± 0.014	53.0 ± 12.4	1064 ± 338
Mutant 110	0.075 ± 0.017	53.9 ± 14.8	731 ± 148
Mutant 244	0.068 ± 0.015	70.6 ± 20.6	1048 ± 242
Mutant 281	0.069 ± 0.015	68.7 ± 4.7	1023 ± 160
Mutant 77[a]	0.048 ± 0.001	153.89 ± 5.84	3193 ± 198
Mutant 110[a]	0.064 ± 0.004	150.25 ±0.35	2344 ± 161
Mutant 244[a]	0.055 ± 0.003	120.76 ± 5.18	2192 ± 65
Mutant 281[a]	0.049 ± 0.003	127.05 ± 3.80	2598 ± 235

Note: Enzyme kinetics performed on synthetic substrate Lys-Pro-Ala-Glu-Phe-Phe(NO_2)-Ala-Leu in 100 mM sodium citrate buffer, pH 2.0, and estimated using nonlinear regression. Kinetic constants were calculated based on three replicates with three determinations per replicate.

[a] Nonglycosylated form of recombinant pepsin.

Source: Ref. 74.

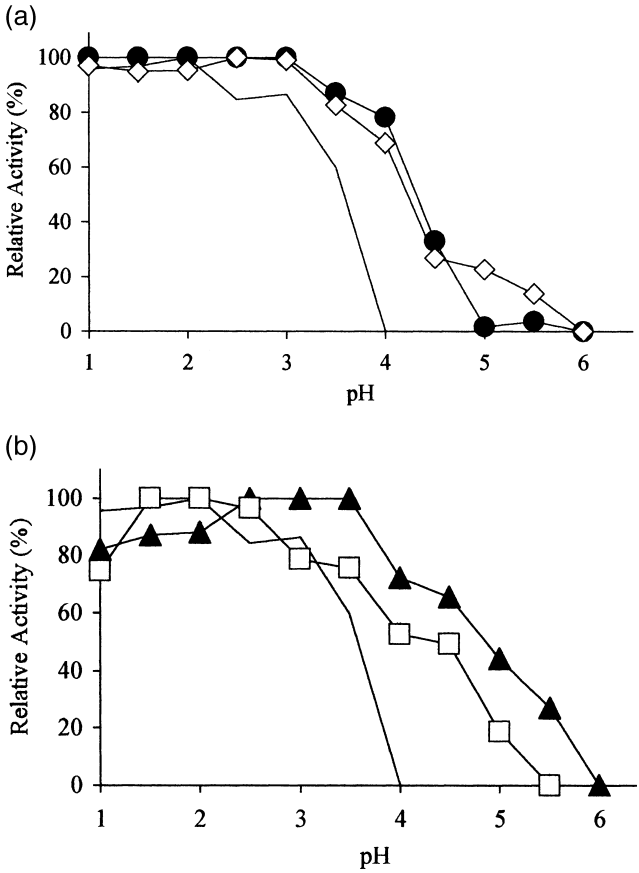


Figure 13 Effect of pH on activation of glycosylated pepsins. Samples were incubated at various pH values for 30 min, quenched at pH 8.0 (overnight, 4°C), and the remaining pepsinogen activated at pH 1.0 for 30 min. The sample was buffered to pH 5.3 and hydrolysis of the synthetic substrate (Lys-Pro-Ala-Glu-Phe-Phe(NO_2)-Ala-Leu) measured at 300 nm. The relative acivity was expressed as a percentage of the highest activity over the pH range examined. Each data point represents the mean of three determinations. (a) Mutant 77 (●), mutant 110 (◇) and wild-type (—); (b) mutant 244 (▲), mutant 281 (□), and wild type (—). (From Ref. 74.)

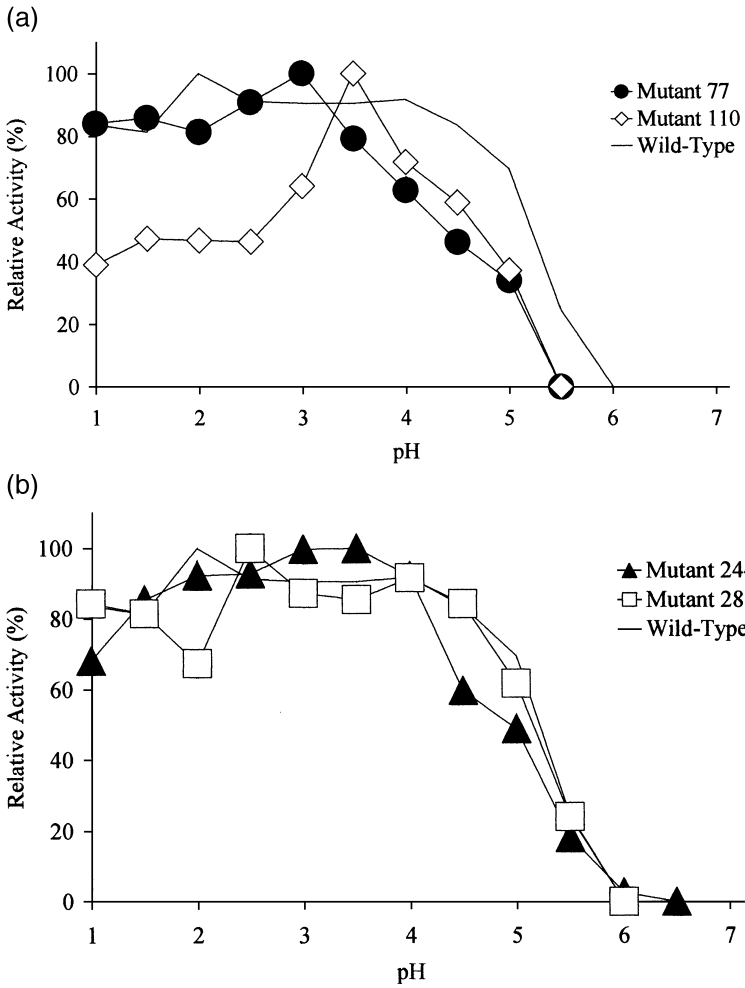


Figure 14 Effect of pH on activity of glycosylated pepsins. Hydrolysis of the synthetic substrate was measured at 300 nm over the various pH values. The relative acitivity was expressed as a percentage of the highest activity over the pH range examined. Each data point represents the mean of three determinations. (From Ref. 74.)

Table 15 Thermodynamic Characteristics of Mutant Pepsins

Recombinant pepsin	Peak A	Peak B
	t_m(°C)	t_m(°C)
Wild type	38.8 ± 1.2	66.2 ± 0.7
Mutant 77	46.1 ± 1.0	69.2 ± 2.3
Mutant 110	49.1 ± 0.7	67.1 ± 0.4
Mutant 244	44.6 ± 1.3	6.9 ± 1.1
Mutant 281	55.5 ± 1.4	68.3 ± 0.5

Note: Samples (0.8 mg per scan) were scanned at a rate of 1.5°C min from 20°C to 80°C. Data represents the mean of three replicates. t_m(°C) = melting point.
Source: Ref. 74.

strate specificity as evidenced by peptide mapping was not altered. Activation of pepsinogen to pepsin occurred between pH 1.0 and 4.0 for wild-type pepsin, whereas the glycosylated recombinants activated over a wider range (i.e., pH 1.0–6.0) (Fig. 13). Glycosylation on the C-terminal domain exhibited pH activity profiles similar to nonglycosylated pepsin, whereas glycosylation on the N-domain resulted in a decrease in activity (Fig. 14). In addition, glycosylation led to a slight increase in structural stability around neutral pH and an increase in thermal stability (Table 15) based on activity and structure (Fig. 15). Overall, glycosylation on the C-domain led to a more global stabilization of the structure which translated into enzymatic stability, whereas on the N-domain, an increase in structural stability had little effect on enzymatic stability. Finally, glycosylation on the flexible loop also appeared to increase the overall structural stability of the protein compared to the wild type. It is postulated that the presence of the carbohydrate residues added rigidity to the protein structure by reducing conformational mobility of the protein, thereby increasing the structural stability.

X. EXAMINATION OF NONMAMMALIAN AP: PLANT

Most of our research to date has examined mammalian aspartic proteinases—in particular, pepsin(ogen); however, we have now begun to examine other aspartic proteinases in attempts to understand the universality among this class of enzymes. Unlike their animal and microbial counterparts, few plant aspartic proteinases have been characterized biochemically. The role of plant APs is highly speculative, with reports suggesting its importance for protein turnover, germination, senescence, insect and microbial protection,

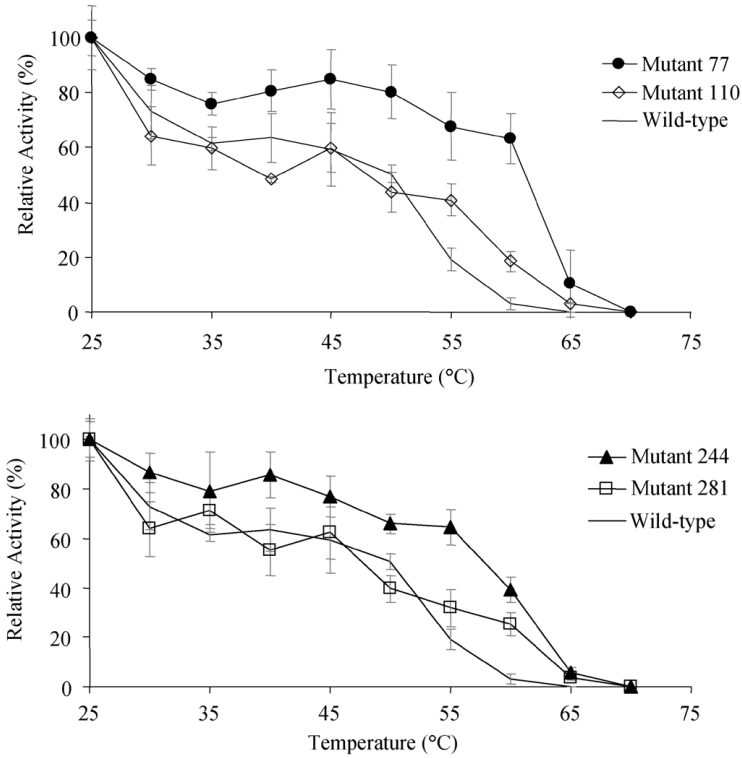


Figure 15 Effect of temperature on stability of glycosylated pepsins. Samples were incubated for 10 min at each temperature and cooled to 4°C and hydrolysis of the synthetic substrate measured at 300 nm. The relative activity was expressed as a percentage of the highest activity over the pH range examined. Each data point represents the mean of three determinations. (From Ref. 74.)

and drought and temperature fluctuations [75]. The specific substrate(s) for these enzymes, however, remain unknown.

Our initial foray into nonmammalian aspartic proteinases examined an aspartic proteinase from alfalfa [76]. Our results indicated that this enzyme had a molecular weight of 33.5 kDa was monomeric and catalyzed the cleavage of a broad spectrum of peptide bonds of hydrophobic amino acids from pH 2.6 to 6.4. The enzyme was inhibited by pepstatin-A (a common inhibitor of many aspartic proteinases). The N-terminal sequence showed 50% and 40% similarity with cyprosin and barley aspartic proteinases, respectively.

Plant aspartic proteinases possess an amino acid segment in their nascent sequence not found in any other AP (i.e., completely novel to the plant king-

dom). The plant specific sequence (PSS) is an internal region of ~100 amino acids (5 α-helices, 3 disulfide linkages, and a N-linked glycosylated moiety), found toward the C-terminal end. Plant APs have been described in both monomeric and heterodimeric forms. Interestingly, although active monomeric plant APs possess the PSS region, heterodimeric plant APs have demonstrated the partial or total loss of the PSS region through autoproteolysis or via other endogenous APs. The reason(s) for the excision of the PSS region remains unclear. Additionally, there appears to be no known reason for the diversity in structure for plant APs; that is, both monocots and dicots can possess either form where type of plant, nor tissue source, appears to be relevant [75]. The novel mosaic form of the plant APs has prompted the crystal structure of a zymogenic plant AP and demonstrated an overall bilobar fold, with a highly α-helical PSS toward the C-terminus [77]. Sequence relatedness has shown the PSS has an ~30% identity to a class of small proteins called saposins (sphingolipid activator molecules). In their pro form, prosaposins are believed to be involved in a targeting function. They have been shown to associate with procathepsin D (human aspartic proteinase) until they reach (as a unit) the heavy lysosome for processing [78]. However, the targeting function of the prosaposin region has not been demonstrated conclusively. Current dogma, however, is that the PSS region, due to its putative targeting function and structural relatedness to the saposin domain, is responsible for targeting plant APs to the vacuole [77,79]. Our hypothesis is that this specialized region possesses a structure–function relationship (SFR) role, which has yet to be characterized, as striking differences in substrate specificity and catalytic efficiencies are found between those APs that possess and remove their internal PSS region [80–82].

XI. RATIONALE FOR RATIONAL REDESIGN RESEARCH

Presently, our research has involved site-directed mutagenesis of specific amino acids to assess their role on the structure–function relationships of pepsin. It has been our intent to expand into more novel and creative protein engineering endeavors. An alternative approach is to use the rational redesign strategy of enzyme structure, which has received less attention in the enzyme engineering than either random and/or designed mutagenesis type experiments. This technique involves the introduction of key residues and/or structural elements that are not available through random or designed mutagenesis events. However, the knowledge of protein structure, mechanism, and a sensitive enzyme assay are critical for successful novel enzyme engineering efforts [83].

In our quest to further elucidate structure–function relationships of plant aspartic proteinases as well as engage in novel protein engineering strategies,

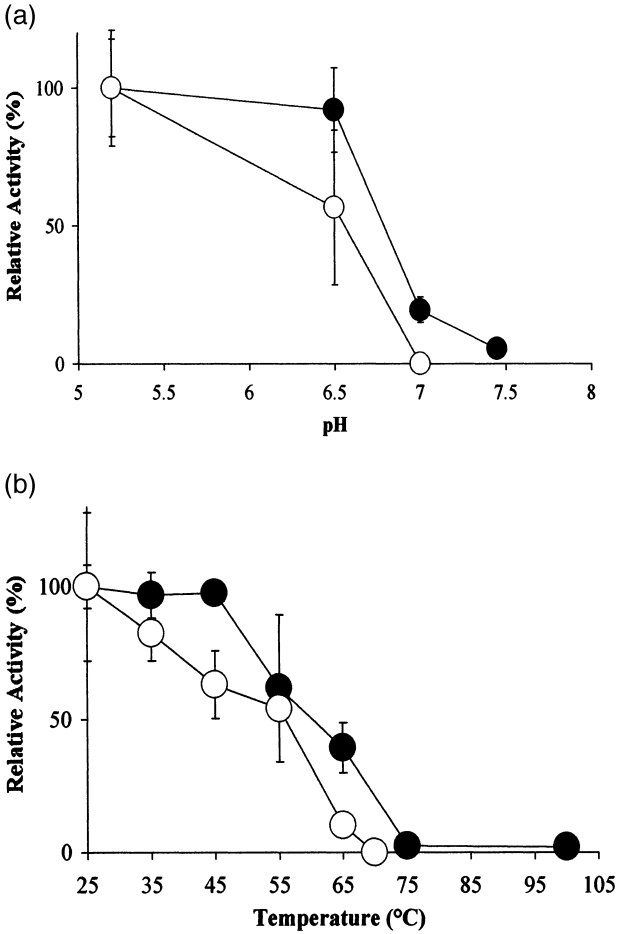


Figure 16 Stability curves of the chimeric enzyme and porcine pepsin. (a) Neutral and alkaline pH stability of the chimeric enzyme and porcine pepsin. Activity was measured after incubation at pH 6.5, 7.0, and 7.5 using 0.4 *M* phosphate buffer for 10 min and adjusted back to pH 5.2, using 3 *M* sodium acetate. Three replicates were performed in triplicate using 0.025 m*M* SS1 at 25°C with the residual mean activity compared to the mean activity at pH 5.2 in 3 *M* sodium acetate (control) with standard deviations plotted. (b) Temperature stability of the chimeric enzyme and porcine pepsin. After incubation at specified temperatures for 10 minutes, the enzyme was returned to 4°C. Three replicate activity measurements were performed in triplicate using 0.025 m*M* SS1, pH 5.2, at 25°C. The mean residual activity was compared to the mean activity at 25°C (control) with standard deviations plotted. (From Ref. 84.)

we undertook a study whereby the PSS domain from the *Arabidopsis thaliana* aspartic proteinase was inserted into porcine pepsinogen at the identical location found in nascent plant aspartic proteinases, to create a chimeric mammalian–plant enzyme [84]. Based on enzymatic activity, this chimeric enzyme demonstrated increases in pH stability above 6 (Fig. 16) and temperature stability above 60°C (Fig. 17) compared to commercial pepsin. Differential scanning calorimetry of the chimeric enzyme illustrated an ~2°C increase in denaturation temperature (T_m) with increases in cooperativity and similar enthalpy values. Kinetic analysis indicated an increase in K_m and decreased k_{cat} compared with pepsin, but with a catalytic efficiency similar to the monomeric plant aspartic proteinase from wheat (Table 16). Using oxidized insulin B-chain, the chimeric enzyme demonstrated more restricted substrate specificity in comparison to commercial pepsin (Fig. 18). This study highlights the use of a chimeric enzyme strategy in order to characterize unique protein domains within enzyme families and, for the first time, a putative structure–function role for the plant specific sequence, as it pertains to plant aspartic proteinases.

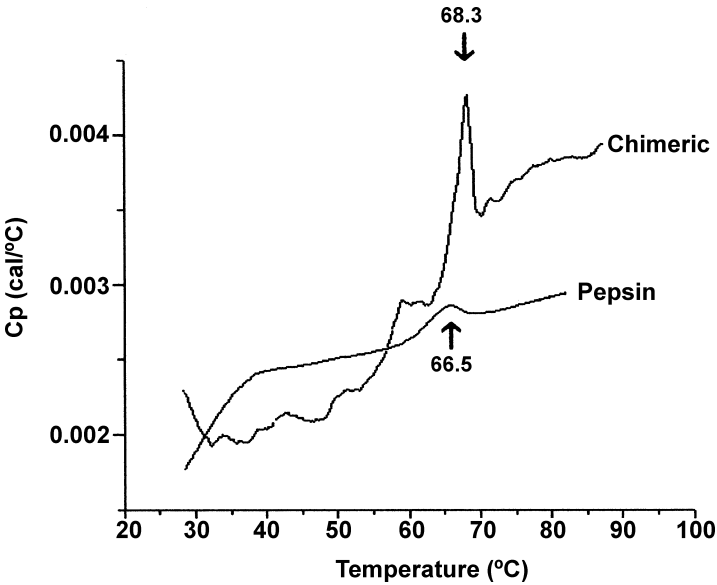


Figure 17 Differential scanning calorimetry (DSC) characterizations of the chimeric enzyme and porcine pepsin. DSC measurements were performed in duplicate using a MicroCal MC-2 differential scanning calorimeter with the T_m values for the chimeric enzyme and pepsin indicated on the figure. (From Ref. 84.)

Table 16 Kinetic Analysis of the Chimeric Aspartic Proteinase and Comparison with Recombinant Pepsin, Commercial Pepsin, and Four Plant Aspartic Proteinases

	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (μM^{-1} s^{-1})
Chimeric enzyme[a]	117 ± 37	8.3 ± 2.9	70.5 ± 7.0
De-*N*-glycosylated Chimeric enzyme[a]	131 ± 9	1.2 ± 0.5	8.9 ± 3.3
r-Pepsin[16,a]	62 ± 3	162 ± 5.8	2600 ± 47
c-Pepsin[16,a]	54 ± 13	106 ± 10.4	2010 ± 260
Cardosin A[41,a]	108 ± 86	55.6 ± 6.1	515 ± 57
Cardosin B[41,a]	110 ± 12	89.4 ± 3.5	808 ± 32
Wheat AP[b]	375 ± NR	25.5 ± NR	68 ± NR
Barely AP[c]	205 ± NR	31 ± NR	149 ± NR

Note: NR = not reported. r-Pepsin denotes recombinant pepsin produced from the *P. pastoris* expression system. c-Pepsin denotes activated, purified commercial pepsinogen (Sigma Chemical Co.).

[a] Kinetics determined using synthetic substrate 1 (KPAEFF(NO_2)AL).

[b] Kinetics determined using synthetic substrate PTEFF(NO_2)RL [30].

[c] Kinetics determined using synthetic substrate KPIEFF(NO_2)RL [22].

Source: Ref. 84.

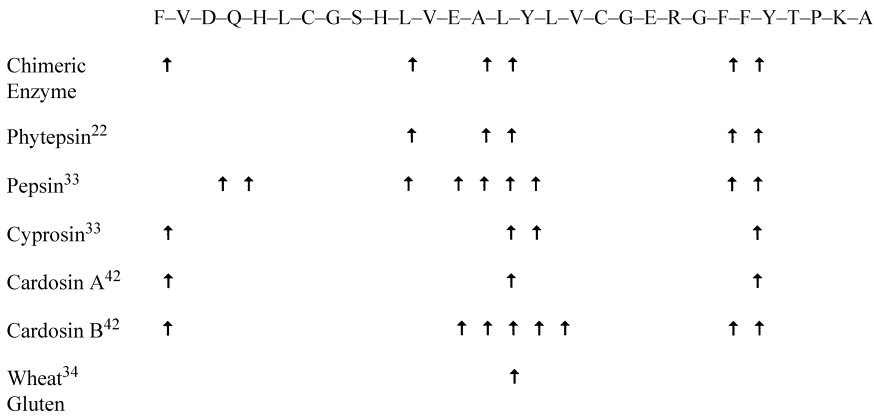


Figure 18 Oxidized insulin B-chain cleavage profile of the chimeric aspartic proteinase in comparison with other aspartic proteinases. Cleavage profile for the chimeric enzyme and pepsin were determined via RP-HPLC with amino acid analysis. Upward arrows represent cleavage sites determined within the oxidized insulin B-chain for the given aspartic proteinase. (From Ref. 84.)

In conclusion, although our studies have been directed toward the structure–function relationships of aspartic proteinases, information will undoubtedly contribute to the structure–function relationships of other food-related proteins. Such information is necessary in order to allow for more efficient utilization of existing and/or new food protein sources and to provide insight into development of strategies for the tailoring of proteins for specific end uses.

ACKNOWLEDGMENTS

The financial support of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged. The assistance of Dr. Hiro Ogawa with the illustrations is also gratefully acknowledged.

REFERENCES

1. Davies, D.R. Annu. Rev. Biophys. Chem. 1990, *19*, 189–215.
2. Creighton, T.E. Prog. Biophys. Mol. Biol. 1978, *33*, 231–298.
3. Georgiou, G.; De Bernardez-Clark, E. In *Protein Folding*; *ACS Series 470*; American Chemical Society: Washington DC, 1991.
4. Kane, J.F.; Hartley, D.L. Trends Biotech. 1998, *X*, 95–101.
5. LaVallie, E.R.; DiBlasio, E.A.; Kovacic, S.; Grant, K.L.; Schendel, P.F.; McCoy, J.M. Bio/Technology 1993, *11*, 187–193.
6. Nilsson, B.; Abrahmsen, L.; Uhlzn, M. EMBO J. 1985, *4*, 1075–1080.
7. Tanaka, T.; Yada, R.Y. Biochem. J. 1996, *315*, 443–446.
8. Al-Janabi, J.; Hartsuck, J.A.; Tang, J. J. Biol. Chem. 1972, *247*, 4628–4632.
9. Bustin, M.; Conway-Jacobs, A. J. Biol. Chem. 1971, *246*, 615–620.
10. Herriott, R.M. J. Gen. Physiol. 1939, *22*, 65–78.
11. McPhie, P. Biochem. Biophys. Res. Commun. 1974, *56*, 789–792.
12. Tanaka, T.; Yada, R.Y. Arch. Biochem. Biophys. 1997, *340*, 355–358.
13. Cregg, J.M.; Cereghino, J.L.; Shi, J.; Higgins, D.R. Mol. Biotechnol. 2000, *16*, 23–52.
14. Yoshimasu, M.A.; Ahn, J.K.; Tanaka, T.; Yada, R.Y. Protein Express. Purif. 2002, *25*, 229–236.
15. Twining, X. Crit. Rev. Biochem. Mol. Biol. 1994, *29*, 315–383.
16. Koelsch, G.; Mares, M.; Metcalf, P.; Fusek, M. FEBS Lett. 1994, *343*, 6–10.
17. Hartsuck, J.A.; Koelsch, G.; Remington, S.J. Proteins 1992, *13*, 1–25.
18. Sielecki, A.R.; Fujinaga, M.; Read, R.J.; James, M.N.G. J. Mol. Biol. 1991, *219*, 671–692.
19. Glick, D.M.; Shalitin, Y.; Hilt, C.R. Biochemistry 1989, *28*, 2626–2630.
20. Richter, C.; Tanaka, T.; Koseki, T.; Yada, R.Y. Eur. J. Biochem. 1999, 746–752.
21. Fetrow, J.S. FASEB J. 1995, *9*, 708–717.
22. Leszczynski, J.F.; Rose, G.D. Science 1986, *234*, 849–855.
23. Rose, G.D.; Gierasch, L.M.; Smith, J.A. Adv. Protein Chem. 1985, *37*, 1–109.
24. Tramontano, A.; Chothia, C.; Lesk, A.M. Proteins 1989, *6*, 382–394.

25. Gerstein, M.; Chothia, C. J. Mol. Biol. 1991, *220*, 133–149.
26. Sampson, N.S.; Knowles, J.R. Biochemistry 1992, *31*, 8488–8494.
27. Kempner, E.S. FEBS Lett. 1993, *326*, 4–10.
28. Sielecki, A.R.; Fedorov, A.A.; Boodhoo, A.; Andreeva, N.S.; James, M.N.G. J. Mol. Biol. 1990, *214*, 143–170.
29. Sali, A.; Veerapandian, B.; Cooper, J.B.; Foundling, S.I.; Hoover, D.J.; Blundell, T.L. EMBO J. 1989, *8*, 2179–2188.
30. Bott, R.R.; Subramanian, E.; Davies, D.R. Biochemistry 1982, *21*, 6956–6962.
31. James, M.N.G.; Sielecki, A.R.; Salituro, F.; Rich, D.H.; Hofmann, T. Proc. Natl. Acad. Sci. USA 1982, *79*, 299–361.
32. Abad-Zapatero, C.; Rydel, T.J.; Nieddhart, D.J.; Luly, J.; Erickson, J. In *Structure and Function of the Aspartic Proteinases*; Dunn, B., Ed.; Plenum Press: New York, 1991; 9–21.
33. Foundling, S.I.; Cooper, J.B.; Watson, F.E.; Cleasby, A.; Pearl, L.H.; Sibanda, B.L.; Hemmings, A.; Wood, S.P.; Blundell, T.L.; Valler, T.L.; Noey, C.G.; Kay, J.; Boger, J.; Dunn, B.M.; Leckie, B.J.; Jones, D.; Atrash, B.; Hallett, A.; Szelke, M. Nature 1987, *327*, 349–352.
34. Andreeva, N.S. Scand. J. Lab. Invest. 1992, *52* (Suppl 210), 31–38.
35. Suzuki, J.; Hamu, Nishiyama, M.; Horinouchi, S.; Beppu, T. Protein Eng. 1989, *2*, 563–569.
36. Beppu, T.; Park, Y.N.; Aikawa, J.; Nishiyama, M.; Horinouchi, S. In *Aspartic Proteinases*; Takahashi, K., Ed.; Plenum Press: New York, 1995; 501–509.
37. Chen, L.; Erickson, J.W.; Rydel, T.J.; Park, C.H.; Neidart, D.; Luly, J.; Abad-Zapatero, C. Acta Crystallogr. 1992, *B48*, 476–488.
38. James, M.N.G.; Sielecki, A.R. Biochemistry 1985, *24*, 3701–3713.
39. Pearl, L.H. FEBS Lett. 1987, *214*, 8–12.
40. Lowther, W.T.; Dunn, B.M. In *Aspartic Proteinases*; Takahashi, K., Ed.; Plenum Press: New York, 1995; 555–558.
41. Tanaka, T.; Teo, K.S.L.; Lamb, K.M.; Harris, L.J.; Yada, R.Y. Protein Peptide Lett. 1998, *5*, 19–26.
42. Okoniewska, M.; Tanaka, T.; Yada, R.Y. Biochem. J. 2000, *349*, 169–177.
43. Creighton, T.E. *Proteins, Structures and Molecular Properties,* 2nd Ed.; W.H. Freeman: New York, 1994.
44. Okoniewska, M.; Tanaka, T.; Yada, R.Y. Protein Eng. 1999, *12*, 55–61.
45. Wells, J.A.; Estell, D.A. In *Proteins: Form and Function*; Bradshaw, R.A., Purton, M., Eds.; Elsevier: New York, 1990; 45–57.
46. Shih, P.; Kirsch, J.F. Protein Sci. 1995, *4*, 2063–2072.
47. Elbein, A.D. Annu. Rev. Biochem. 1987, *56*, 497–534.
48. Ottesen, M.; Rickert, W. Compt. Rend. Trav. Lab. Carlsberg 1970, *37*, 301–325.
49. Garnot, P. Int. Dairy Fed. Bull. 1985, *194*, 2–7.
50. Bohak, Z. J. Biocl. Chem. 1969, *244*, 4638–4648.
51. Fruton, J.S. Adv. Enzymol. 1971, *33*, 401–443.
52. James, M.N.G.; Sielecki, A.R. Nature 1986, *319*, 33–38.
53. Lin, X.L.; Wong, R.N.; Tang, J. J. Biol. Chem. 1989, *264*, 4482–4489.
54. Lin, X.L.; Loy, J.A.; Sussman, F.; Tang, J. Protein Sci. 1993, *2*, 1383–1390.
55. Lee, A.Y.; Gulnik, S.V.; Erickson, J.W. Nature. Struct. Biol. 1998, *5*, 866–871.
56. Tanaka, T.; Yada, R.Y. Protein Eng. 2001, *14*, 669–674.

57. Bryksa, B.C.; Tanaka, T.; Yada, R.Y. Biochemistry 2003, 13331–13338.
58. Cooper, J.B.; Khan, G.; Taylor, G.; Tickle, I.J.; Blundell, T.L. J. Mol. Biol. 1990, *214*, 199–222.
59. Rademacher, T.W.; Parekh, R.B.; Dwek, R.A. Annu. Rev. Biochem. 1988, *57*, 785–938.
60. Paulson, J.C. In *Proteins: Form and Function*; Bradshaw, R.A., Purton, M., Eds.; Elsevier: Cambridge, 1990; 207–217.
61. Matthews, C.R. Annu. Rev. Biochem. 1993, *62*, 653–683.
62. Imperiali, B.; Rickert, K.W. Proc. Natl. Acad. Sci. USA 1995, *92*, 97–101.
63. Wang, C.; Eufemi, M.; Turano, C.; Giartosio, A. Biochemistry 1996, *35*, 7299–7307.
64. Hasagawa, M.; Hidaka, Y.; Wada, A.; Hirayama, T.; Shimonishi, Y. Eur. J. Biochem. 1999, *263*, 338–345.
65. Letourneur, O.; Gervasi, G.; Gala, S.; Pages, J.; Watelet, B.; Jolivet, M. Biotechnol. Appl. Biochem. 2001, *33*, 35–45.
66. Helenius, A. Mol. Biol. Cell. 1994, *5*, 253–265.
67. Haraguchi, M.; Yamashiro, S.; Furukawa, K.; Takamit\ya, K.; Shiku, H.; Furukawa, K. Biochem. J. 1995, *312*, 272–280.
68. McGinnes, L.W.; Morrison, T.G. Virology 1995, *212*, 398–410.
69. Newrzella, D.; Stoffel, W. J. Biol. Chem. 1996, *271*, 32,089–32,095.
70. Levy, O.; Deal Viega, A.; Ginter, C.S.; Riedel, C.; Dai, G.; Carrasco, N. J. Biol. Chem. 1998, *273*, 22,657–22,663.
71. Miele, R.G.; Prorok, M.; Costa, V.A.; Castellino, F.J. J. Biol. Chem. 1999, *274*, 7769–7776.
72. Nakamura, S.; Takasaki, H.; Kobayashi, K.; Kato, A. J. Biol. Chem. 1993, *268*, 12,706–12,712.
73. Shu, Y.; Maki, S.; Nakamura, S.; Kato, A. J. Agric. Food Chem. 1998, *46*, 2433–2438.
74. Yoshimasu, M.A.; Tanaka, T.; Ahn, J-K.; Yada, R.Y. Glycobiology 2004. *in press*.
75. Mutlu, A.; Gal, S. Physiol. Plant 1999, *105*, 569–576.
76. Payie, K.G.; Weadge, J.T.; Tanaka, T.; Yada, R.Y. Biotechnol. Lett. 2000, *22*, 1515–1520.
77. Kervinen, J.; Tobin, G.J.; Costa, J.; Waugh, D.S.; Wlodawer, A.; Zdanov. EMBO J 1999, *18*, 3947–3955.
78. Zhu, Y.; Conner, G.E. J. Biol. Chem. 1994, *269*, 3846–3851.
79. Egas, C.; Lavoura, N.; Resende, R.; Brito, R.M.M.; Pires, E.; Pedrosa de Lima, M.C.; Faro, C. J. Biol. Chem. 2000, *275*, 38,190–38,196.
80. Verissimo, P.; Faro, C.; Moir, A.J.; Lin, Y.; Tang, J.; Pires, E. Eur. J. Biochem. 1996, *235*, 762–768.
81. Galleschim, L.; Feicioli, F. Plant Sci. 1994, *98*, 15–24.
82. Kervinen, J.; Sarkinen, P.; Kalkkinen, N.; Mikola, N.; Saarma, M. Phytochemistry 1993, *32*, 799–803.
83. Chen, R. Trends Biotechnol. 2001, *19*, 13–14.
84. Payie, K.; Tanaka, T.; Gal, S.; Yada, R.Y. Biochem. J. 2003, *372*, 671–678.

10

Computer Simulation of Soft Mesoscopic Systems Using Dissipative Particle Dynamics

DAVID A. PINK St. Francis Xavier University, Antigonish, Nova Scotia, Canada

I. INTRODUCTION

Many systems of interest to the food and health industries are mesoscopic: that is, their components give rise to interacting phenomena encompassing a wide range of spatial scales and timescales which are, respectively, much larger and much longer than the characteristic spatial scales and timescales of the components themselves. Some examples are hydrogels with embedded proteins or other macromolecules, polymer mixtures, and liposomes and vesicles as delivery systems. In order to understand observations made on such systems, it is important to have mathematical models which can be used to make predictions. The operator which describes the total energy of a system, the Hamiltonian operator, can be written down, but a number of problems immediately arise. The first is whether we make use of classical or semiclassical models of atomic phenomena or whether we try to include quantum effects. At present, it is unrealistic to attempt to address quantum effects in mesoscopic systems and here we restrict ourselves to the first approach. Again, two decisions must be made. What are the ranges of length scales and timescales that the model must describe? To require a description of molecular phenomena on the smalles length scale of, say, 10^{-2} nm or less requires atomic models in which details of the atomic electron densities, $|\psi|^2$ are utilized. This generally takes the form of representing atomic constituents of molecules by shapes reflecting their van der Waals radii. These interact among themselves via various potential of which the 6–12 van der Waals potential is the simplest. If, however, one is satisfied with a larger scale, then some form of

"minimal model" might be used. In this case, the structures employed to represent molecular shapes do not necessarily reflect the details of electron densities.

Having created a mathematical model, one must now use it to calculate static quantities such as polymer radii of gyration, static structure functions, and other quantities that can be measured. One can also calculate dynamic quantities such as diffusion coefficients, dynamic structure functions, relaxation times, and dynamic viscosities. In order to do this, we must effectively solve equations of motion. It is possible that the equations of motion are sufficiently tractable that analytic methods are adequate to obtain solutions to them. In the case of complex mesoscopic systems such a satisfactory outcome is unlikely and one must resort to numerical methods—computer simulation.

If we are interested only in static quantities, then there are basically two computer simulation techniques that we can use: Monte Carlo (MC) methods or molecular dynamics (MD). The first employs the Metropolis algorithm [1], a computer simulation technique that eventually generates a sequence of states (configurations) of the system that reflect thermal equilibrium at temperature T. This sequence of states can be used to calculate equilibrium thermodynamic quantities. In practice, the calculations proceed while the states are being generated. The second, MD, technique calculates both static and dynamic averages simultaneously. It involves solving Newton's equations [2] for the dynamics of the semiclassical model that we are using to represent the mesoscopic system.

$$\frac{d}{dt}\vec{r}_j = \vec{v}_j \tag{1}$$

$$m_j \frac{d}{dt}\vec{v}_j = \vec{f}_j \tag{2}$$

where the set $\{\vec{r}_j, \vec{v}_j, \vec{f}_j\}$ are the position and velocity of the jth member of the system and the force acting on it. Because of the complexity of the models, it is necessary to solve these equations numerically using methods appropriate for coupled differential equations. One of these is described below. The MD technique simultaneously calculates both dynamic and static properties. One question that arises is "can the MC method calculate dynamical quantities, such as a diffusion coefficient?" and the answer is "possibly." It has been argued that if thermal fluctuations dominate the equations of motion and if the sequence of "equilibrium" configurations generated are "sufficiently close" in the phase space of the system, then the sequence of states represents the time evolution of the system through phase space and each MC step represents some elapsed time.

II. MONTE CARLO METHOD

The Metropolis algorithm generates configurations of a system in a sequence of MC steps in the following way [3,4]. Let the system, at MC step M and temperature T, be in a state (configuration) possessing energy E_p. Randomly choose another state, q, possessing energy E_q. Define $\Delta E = E_q - E_p$ and $\beta = 1/k_B T$, where k_B is Boltzmann's constant. If $\Delta E < 0$, then the system is placed in state q at MC step $M + 1$. If $\Delta E > 0$, then choose a random number R with $0 \leq R < 1$. If $R \leq e^{-\beta \Delta E}$, then the system is placed in state q at MC step $M + 1$. Otherwise, the system remains in the state p at MC step $M + 1$. If the states p and q are "sufficiently close to each other" in the phase space of the system and if fluctuation forces dominate the dynamics, then it is possible that this technique approximates the time development of the system. There is, however, no guarantee that the magnitude of the time step is independent of the path that the system follows through the phase space. It has been proven that this procedure will eventually generate a sequence of state $\{p, q, \ldots\}$ characteristic of thermodynamic equilibrium. The system is first set up in any convenient initial state. The MC procedure is then carried out using a sufficiently large number of MC steps so that the system relaxes. One then continues generating configurations, at the same time accumulating data describing quantities of interest (radius of gyration, position). After M such additional MC steps, one obtains an average value for the quantities of interest by dividing by M. The average thus obtained is an equilibrium thermal average for the quantities of interest. In practice, one chooses as an initial configuration any state of the system because the resulting averages must be independent of it. However, there are reservations to this last comment: If a system is in a coexistence regime near a discontinuous phase transition, then the resulting equilibrium thermal averages can, correctly, depend on the starting state of the system.

III. DISSIPATIVE PARTICLE DYNAMICS

At the present time, atomic models are unsuited for studying mesoscopic systems: Both the spatial scales and timescales are far too large to be able to carry out a simulation in any reasonable length of time. Dissipative particle dynamics (DPD) is a modeling technique that utilizes a model with specific properties in order to achieve a well-specified end and then uses MD techniques to obtain static and dynamic information. It is one of new models that were developed in the last decade which include lattice gas cellular automata [5–7]. Lattice gas models, however, are inadequate for modeling complex macromolecular systems such as polymer solutions and bacterial surfaces [8].

Dissipative particle dynamics was invented in order to correctly describe the hydrodynamics of fluids such as water on a mesoscopic scale (macromo-

lecular to micron or submicron), yet, with sufficiently simple interactions, that computer simulation of the model could be relatively fast and efficient (compared to MD of atomic models) [9,10]. Here, "correctly describe" meant that where a classical continuum model of fluid flow can be compared to the results of a simulation using DPD, the latter are in accord with the results of the dynamical equation which describes the former, the Navier–Stokes equation.

Dissipative particle dynamics represents the components of a system as a set of objects—"packets of material" [11]. The components can be clusters of atoms [e.g., $N(CH_3)_3$ or CH_2] or larger clusters [e.g., $(CH_2)_n$] or a sequence of sugar groups, or larger structures. The objects which represent these components are point "particles" which interact among themselves and with externally applied fields. Each pair of particle i and j experiences three kinds of force describing their interactions with each other: a conservative force, $\overrightarrow{F^C}(i,j)$, a dissipative force, $\overrightarrow{F^D}(i,j)$, and a random force, $\overrightarrow{F^R}(i,j)$. If the particles have position and velocity vectors, $\vec{r}_i$, $\vec{r}_j$, $\vec{v}_i$, $\vec{v}_j$ and if the ranges of their interaction is r_{i0}, r_{j0} and if we define $r_{ij0} = r_{i0} + r_{j0}$ (i.e., the three forces are zero if the distance between the particles, $r_{ij} = |\vec{r}_i - \vec{r}_j|$, is greater than r_{ij0}), then for $r_{ij} < r_{ij0}$,

$$\overrightarrow{F^C}(i,j) = a_{ij}\left(1 - \frac{r_{ij}}{r_{ij0}}\right)\widehat{r_{ij}} \tag{3}$$

$$\overrightarrow{F^D}(i,j) = -\gamma_{ij}\left(1 - \frac{r_{ij}}{r_{ij0}}\right)^2\left(\widehat{r_{ij}} \cdot \overrightarrow{v_{ij}}\right)\widehat{r_{ij}} \tag{4}$$

$$\overrightarrow{F^R}(i,j) = \sqrt{2\gamma_{ij}k_BT}\left(1 - \frac{r_{ij}}{r_{ij0}}\right)\zeta_{ij}\,\widehat{r_{ij}} \tag{5}$$

and $\overrightarrow{F^C}(i,j) = \overrightarrow{F^D}(i,j) = \overrightarrow{F^R}(i,j) = 0$ otherwise. Here, $\widehat{r_{ij}} = (\vec{r}_i - \vec{r}_i)/r_{ij}$, a unit vector, $\vec{v}_{ij} = \vec{v}_l - \vec{v}_j$, a_{ij} and γ_{ij} are the strengths of the conservative and dissipative forces, respectively, and the strength of the random force, $\sqrt{2\gamma_{ij}k_BT}$, is obtained from that of the dissipative force from the fluctuation–dissipation theorem [10]. The factor ς_{if} in the random force are Gaussian random numbers, uncorrelated for different pairs of particles and for different times, possessing zero mean and unit standard deviation. Although we have described only interactions, it is clear that we can apply a force to particle j alone by including a force $F_j\hat{u}$, where F_j is the magnitude of the force acting only on particle j and $\hat{u}$ is a unit vector in the direction of the force. The total force acting on particle j is then

$$\overrightarrow{F_j} = \sum_i \left(\overrightarrow{F^C}(i,j) + \overrightarrow{F^D}(i,j) + \overrightarrow{F^R}(i,j)\right) + F_i\hat{u} \tag{6}$$

It should be clear that if the system contains different species (e.g., water, ions, different kinds of polymers labeled by superscripts k and l), then the force strengths, force ranges, and the forces become, for example, a_{ij}^{kl}, γ_{ij}^{kl}, r_{i0}^{k}, and $\vec{F}^{Dkl}(i, j)$, and so on.

If some of the particles are to be physically connected to others, thus representing a polymer or an aggregate, this can be achieved by connecting them via a massles spring that does not interact with any other particle in the system. The force due to a harmonic spring, representing particles interacting with particle j, can be obtained from a harmonic potential (e.g., Refs. 11 and 12),

$$U^{H}(i, j) = \frac{1}{2} K^{S}(l_0 - r_{ij})^2 \tag{7}$$

or, in the case of a FENE (finitely extensible nonlinear elastic) spring from the logarithmic potential,

$$U^{\mathrm{FENE}}(i, j) = -\frac{1}{2} K^{S} l_0^2 \ln\left[1 - \left(\frac{r_{ij}}{l_0}\right)^2\right] \tag{8}$$

If a semiflexible polymer is required, a force, F_i^{SF}, can be defined at particle i in a polymer chain acting in a direction so as to drive the angle between the two bonds connecting particle i to its nearest neighbors, toward a value of π (e.g., Ref. 13). In this case, two forces, $-1/2\ \vec{F_i^{SF}}$, can act on each of the two neighbors of particle i, $i + 1$, and $i - 1$. This force will work so as to keep the polymer straight. It can be seen from the description of the springs which hold a polymer together that two polymers can pass through each other and that a polymer can pass through itself. If particles i and j are nearest neighbors in a polymer connected by a harmonic or a FENE spring, one should ensure that the radius of their interaction, r_{ij0}, the magnitude of the conservative force, a_{ij}, the spring constant, K^S, and the unstretched length, l_0, are all chosen so that the ability of the polymer chain to pass through itself, or another polymer to pass through it, is effectively eliminated.

In the forces described here, it is clear that Newton's laws are satisfied (for every force there is an equal and opposite force) so that linear momentum will be conserved. There is also the question of whether the requirement of detailed balance is satisfied—that, because Newton's laws are invariant under time reversal ($t \to -t$), the probability of particle i moving from $\vec{r}_i$ to $\vec{r}_i + \overrightarrow{\Delta r_i}$ is equal to the probability of it moving from $\vec{r}_i + \overrightarrow{\Delta r_i}$ to $\vec{r}_i$.

IV. VELOCITY–VERLET ALGORITHM

One algorithm to simulate the time development of the system using Newton's equations is the velocity–Verlet algorithm [2]. At every time step from

t to $t + \delta t$, the position, velocity and acceleration vectors $\{\vec{r}_j, \vec{v}_j, \vec{a}_j\}$, of the $j-th$ mass, m_j, are updated from those at the earlier time in the following way. Given $\vec{r}_j(t)$ and $\vec{v}_j(t)$, calculate $\vec{a}_j(t) = \vec{f}_j(t)/m_j$ where $\vec{f}_j(t)$ is the total force on mass m_j at time t. Then,

$$\vec{r}_j(t+\delta t) = \vec{r}_j(t) + \delta t \vec{v}_j\,(t) + \frac{1}{2}\delta t^2\,\vec{a}_j(t) \tag{9}$$

Define

$$\vec{v}_j\left(t+\frac{1}{2}\delta t\right) = \vec{v}_j(t) + \frac{1}{2}\delta t\,\vec{a}_j(t) \tag{10}$$

Using

$$\vec{a}_j(t+\delta t) = \frac{\vec{f}_j(t+\delta t)}{m_j} \tag{11}$$

calculate

$$\vec{v}_j(t+\delta t) = \vec{v}_j\left(t+\frac{1}{2}\delta t\right) + \frac{1}{2}\delta t\vec{a}_j(t+\delta t) \tag{12}$$

We now have $\vec{r}_j(t+\delta t)$ and $\vec{v}_j(t+\delta t)$. Instead of calculating the velocity at $t + 1/2\delta t$, it can be calculated with the factor of $1/2$ replaced by λ on both sides of Eq. (10). Den Otter and Clarke [14] studied different choices for λ. In almost all of the work described next, the value of $\lambda = 0.5$ was used.

V. FUNDAMENTALS AND APPLICATIONS OF DPDs

A fundamental question is: What is the theoretical underpinning for DPD? As originally presented and used, and as described earlier, it has not been deduced rigorously from a more microscopic model and although it has been shown [15,16] that, for a one-component DPD fluid, the Navier–Stokes equations are obtained in the large scale limit, we can ask whether its use in modeling the complex systems outlined below is justified.

Flekkøy et al. [17,18] derived a DPD-like model by a coarse-graining procedure applied to atomic model MD. They obtained dissipative and random forces from viscous and heat conduction effects in atomic MD. However, their coarse-graining procedure results in, not particles with spherically symmetric properties (as earlier), but cells on a Voronoi lattice defined by the particles and which possess variable sizes and masses. Although this model might be a better representation of a fluid than the one described here, there are substantial practical difficulties in implementing it because the Voronoi structure associated with a given particle changes shape from one time to the next. Serrano et al. [19] have studied the mesoscopic dynamics of Voronoi fluid particles. There are subroutines for this freely available on the Web. Other work on detailed balance is found in Refs. 20 and 21.

Español [22] has reviewed DPD, summarizing the difficulties involving establishing the magnitude of the friction coefficient, γ_{ij} [16], the fact that the conservative forces fix the thermodynmaic behavior of the system [23], and the ambiguity in the length scales and timescales. He proposes expanding the properties of the particle beyond that described here so that they possess additional internal variables and, accordingly, be considered as thermodynamic subsystems. He used the GENERIC formalism [24–26] to write a stochastic differential equation from which he derives the DPD model described here with thermal fluctiuations included in a consistent way. Pagonabarraga et al. [27] pointed out that the temperature calculated in a DPD simulation, $3k_BT = (1/N)\langle\Sigma_j m_j v_j^2\rangle$, depends on the value of δt and also differs from the value calculated using the fluctuation–dissipation theorem. They introduced a new algorithm which goes far to addressing the problems and preserves a microscopic form of detailed balancing.

Other work on the fundamentals of DPD involves obtaining Fokker–Planck equations, developing smoothed DPD, hydrodynamics, transport coefficients and heat conduction in DPD [14,28–36].

In addition to these, work has been done on boundary conditions [37]. This is an important topic because eventually a model must be made as to how DPD particles interact with a surface that is attached to the "rest of the universe." At some point in the heirarchy of particles in a bulk → particles adjacent to a surface → particles on the surface → surface connections to the rest of the universe, boundary conditions must be imposed in order to terminate the heirarchy of interactions while for example, conserving linear momentum. Collective effects (e.g., velocity autocorrelation function) have been considered in [Refs. 38–40]. Hydrodynamics, immiscibility, spinodal decomposition, and viscoelastic flow has been modeled by in Refs. 30 and 41–43, as well as computer simulations of non-newtonian fluids in Ref. 44. Masters and Warren [45] and Trofimov et al. [46] studied kinetic theory and DPD [45,46]. Finally, Willemsen et al. [47] studied a combined DPD–MC technique.

The inclusion of electrostatic effects offers a special problem to DPD which depends on the finite range of the forces to realize a tractable model. Without invoking a Debye screening length, unscreened electrostatic interactions are infinite-range interactions. Two options offer themselves: explicit summing over all charges or coupling all particles to a local electric field for which Poisson's equations are separately solved. In the latter case, it is advantageous to do the calculation in Fourier space [48]. One advantage of this approach is that summation over image charges need not be done. Given the importance of electrostatic effects in mesoscopic systems, this article marks a significant advance.

Dissipative particle dynamics has been used to model the dynamics and rheology of polymer solutions and the effect of solvent quality on the dynamics of polymers [11,49–51], polymer liquid–liquid demixing in polymer

solutions and (liner and branched) polymer–solvent phase separation [52,53], grafted polymer chains between two walls and grafted polymer brushes under shear [54,55], and polymer scaling laws obtained from the use of DPD [56]. Gibson et al. have modeled particle adsorption onto polymer coatings and Elliott and Windle studied filler particles in polymer composites [57,58].

Whittle and Dickinson modeled a colloidal system of spherical particles and included an exponential interaction that could represent either a shielded coulomb potential or steric forces [59]. DPD has resulted in advances in the modeling of self-assembly of soft complex structures. Atomic, or even minimal, models and the use of MD have been inadequate to exhibit ramifications of self-assembly of, for example, lipid molecules into bilayers or vesicles because of, computer-imposed, restrictions on the size of the system and the time required for self-assembly to be manifested. With the advent of DPD, this problem has been bypassed. Work has been done on the formation of lipid bilayer phases and spontaneous vesicle formation [13,60–63].

VI. AN EXAMPLE. A LARGE SPHERICAL BEAD MOVING UNDER A CONSTANT APPLIED FORCE THROUGH AN AQUEOUS SOLUTION OF LINEAR, SEMIFLEXIBLE POLYMERS

This example is of interest because it is a model of the use of the "magnetic tweezers" concept: a paramagnetic bead in an inhomogeneous magnetic field [64–67] moving through a polymeric solution. This technique has been used to study the viscoelastic properties of the solution because the force itself does not perturb the solution. Assume that the magnetic field is along the y axis and also inhomogeneous along the y axis, $\overrightarrow{H(y)} = H(y)\hat{y}$, where $\hat{y}$ is a unit vector. Then, if the induced bead magnetic dipole moment is $\overrightarrow{\mu} = \chi H(y)\hat{y}$ (we have ignored corrections due to bead shape), the force on the bead is

$$\overrightarrow{F} = 2\chi H(y)\ \frac{\partial H(y)}{\partial y}\hat{y} \tag{13}$$

If the functional dependence of $H(y)$ upon y is only slowly varying along the y axis, this force is approximately constant. In what follows we assume that this is true.

Here, we modeled an aqueous solution containing 20 linear polymers and 1 bead so that we have three different species: bead (b), water (w), and monomer of a polymer (m) (see above). Accordingly, the superscripts k, l are b, w, m. Each polymer was made up of 20 monomers connected by harmonic springs. The size of the simulation volume was $40 \times 160 \times 40$ units in the x, y, and z directions, respectively (with the applied magnetic field along the y axis), with

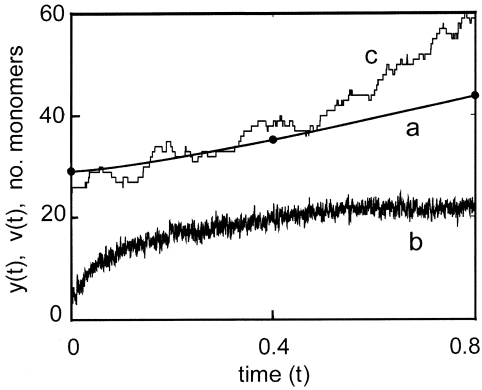


Figure 1 Bead position (a) and velocity (b) along the y axis and number of monomers interacting with the bead (c) as a function of time. The solid circles indicate the times for the configurations of Fig. 3.

periodic boundary conditions. We chose the parameters to be $r_0^b = 10.0$, $r_{i0}^w = 1.0$, $r_0^m = 1.5$, $a_{ij}^{mm} = 10.0$, $\gamma_{if}^{mm} = 1.5$, $a_{ij}^{wm} = 5.0$, $\gamma_{ij}^{wm} = 1.0$, $a_{ij}^{ww} = 5.0$, $\gamma_{ij}^{ww} = 1.0$, $a_{ij}^{bm} = 10.0$, $\gamma_{ij}^{bm} = 1.5$, $a_{ij}^{bw} = 5.0$, $\gamma_{ij}^{bw} = 1.0$. We chose the mass of the bead to be 1.0 and the mass of each monomer and water particle to be 0.06. The unit length is the diameter of the water particle, $2r_{i0}^w = 2.0$, so that the total volume, V, in these units is 3.2×10^4. We chose the density of water particles per unit volume—number of water particles in V [13]—to be $\rho \approx 3$ (a total of 1.85×10^5 water particles) as used by others (e.g., Ref. 13). Two adjacent monomers along a given polymer chain were connected by a har-

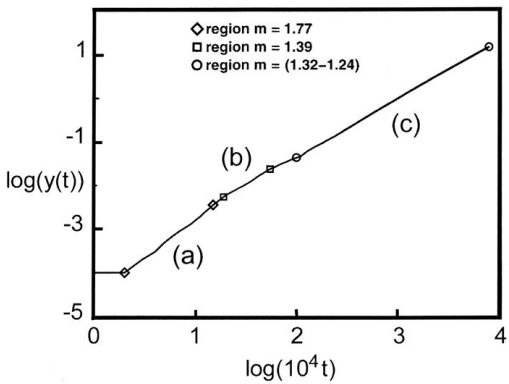


Figure 2 Log [$y(t)$] versus log(10^4 t). Three regions are identified by symbols with the average slopes, m, given at the top of the figure.

monic spring and we chose the spring constant to be 35.0 with $l_0 = 2.5$ Polymers were made semiflexible by conservative forces. If we label the bonds between successive monomers, $i-1$, i, and $i+1$, on a given polymer as $b_{i-1,i}$ and $b_{i,i+1}$, then a force, $\overrightarrow{F_{\theta l}}$, acted on monomer i in a direction bisecting the smaller angle, $\theta < \pi$, between the $b_{i-1,i}$ and $b_{i,i+1}$, linear in $\pi - \theta$, and directed so as to increase this angle. At the same time, forces, $-\overrightarrow{F_{\theta l}}/2$, acted on monomers $i-1$ and $i+1$. The magnitude of $\overrightarrow{F_{\beta i}}$ was chosen so that the

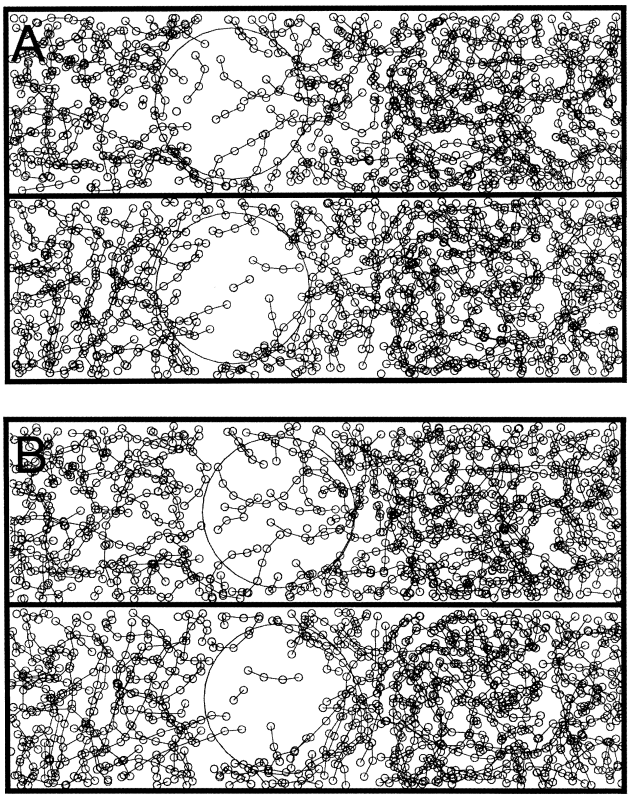


Figure 3 Instantaneous snapshot of the bead (large circle) and monomers (small circles) at three values of t. Monomers which are nearest neighbors on the same polymer have been connected by solid lines. The diameter of the bead is 20 units, and in order to see the bead and the monomers which are interacting with it, we show a slice of the simulation volume with dimensions $23 \times 80 \times 23$ with the bead moving to the right (positive y axis) along the approximate center of the slice. (A) Initial distribution, $t = 0$; (B) at 4000 time steps, $t = 0.400$; (C) at 8000 time steps, $t = 0.8$. In each of these, we show views of the x–y plane (upper) and the z–y plane (lower).

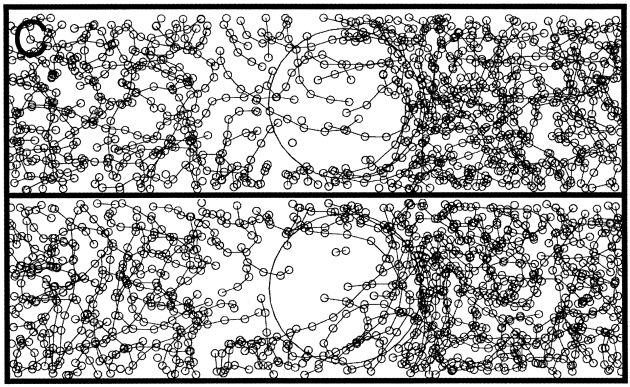


Figure 3 Continued.

persistence length of the polymers was approximately five units, which yielded an average internal angle between successive bonds of $\theta \approx 145°$. Finally, we chose the force due to the inhomogeneous magnetic field to have a magnitude of 5×10^3, chose $k_B T = 1.0$ and $\lambda = 0.5$, and the time step to be $\delta t = 10^{-4}$.

We carried out a simulation run for an elapsed time of $t = 0.8$ (8000 time steps) and some results are shown in Figs. 1 and 2. Figure 1 shows the distance moved by the bead, $y(t)$, as a function of t, its speed along the y axis, $v(t)$, and the total number of monomers interacting with it. The last named shows why the acceleration has gone to zero: The polymer density ahead of the moving bead has increased over the elapsed time. Figure 2 shows a plot of $\log[y(t)]$ as a function of $\log(10^4 t)$. The initial acceleration regime is identified by a slope of $m \approx 1.8$. This is followed by a small region in which the slope is $m \approx 1.4$ followed by a broad region in which the slope appears to change continuously from $m = 1.32$ to $m \approx 1.24$. Figures 3A–3C show a slice of the simulation volume, $23 \times 80 \times 23$, with the bead (diameter $= 20.0$) in the center and space on either side of it to show those polymers which are interacting with it (in some cases, one sees that only short segments of polymers are interacting with the bead). The water particles are not shown. Instantaneous configurations for three times are shown: $t = 0$, $t = 0.4$, and $t = 0.8$; they are related to the data of Fig. 1 by filled circles on the curve showing the distance, $y(t)$, moved by the bead. In the interests of clarity, the radius of the monomers is plotted as 0.5 and does not reflect the monomer force range value of $r_{i0}^m = 1.5$. Monomers connected by a solid line belong to the same polymer. In Figs. 3B and 3C, one can see the buildup of polymers ahead of the bead, the flow of the polymers around the bead and the absence of polymers immediately behind the bead.

VII. CONCLUSIONS

In the decade since it was invented, dissipative particle dynamics has thrown new light on mesoscopic physics, such as self-assembly and polymer dynamics. It is a true breakthrough in modeling and simulating complex materials because it can handle large systems for long times. In this time, it has been studied with a view to putting it on a firmer theoretical foundation. Although the Voronoi lattice realization of DPD is perhaps on firmest theoretical ground, nontheless the simplicity of the original formulation in terms of soft, spherically symmeteric particles all of equal size (for a given species) has been the preferred vehicle for applications. DPD has provided challenges to mathematical theorists and has thrown light on the processes that yield (e.g., self-assembly of lipid bilayers and vesicles). Because of its ease of implementation, it promises to be a major new theoretical tool to model mesoscopic systems.

ACKNOWLEDGMENTS

It is a pleasure to thank Bonnie Quinn for her skillful code-writing, her paranoia concerning our results, and her critical approach. I have been fortunate to have Alexei Boulbitch, Erich Sackmann, Nikita Ter-Oganessian (Technical University of Munich), and Gillian Ryan (St. Francis Xavier University) as colleagues to discuss DPD. Thanks to Reinhard Lipowsky, Julian Shillcock, and Jeremy Pencer for my enjoyable visits to the Max-Planck-Institut for Colloid and Interface Research (Potsdam) where JS and JP introduced me to DPD. This work was supported by NSERC of Canada and the MNG "Structure-Function of Food Related Biopolymers."

REFERENCES

1. Metropolis, N.; Rosenbluth, A.; Rosenbluth, M.; Teller, A.; Teller, E. J. Chem. Phys. 1953, *21*, 1087.
2. Allen, M.P.; Tildesley, D.J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.
3. Binder, K., Ed.; *Monte Carlo Methods in Statistical Physics*, *II*; Springer-Verlag: Heidelberg, 1984.
4. Mouritsen, O.G. *Computer Studies of Phase Transitions and Critical Phenomena*; Springer-Verlag: Heidelberg, 1984.
5. Doolen, G.D., Ed.; *Lattice Gas Methods for Partial Differential Equations*; Addison-Wesley Publishing Company: Reading, MA, 1990.
6. Rothman, D.H.; Zaleski, S. Rev. Mod. Phys. 1994, *66*, 1417.
7. Benzi, R.; Succi, S.; Vergassola, M. Phys. Rep. 1992, *222*, 145.

8. Pagonabarraga, I.; Hagen, M.H.J.; Frenkel, D. Europhys. Lett. 1998, *42*, 377.
9. Hoogebrugge, P.J.; Koelman, J.M.V.A. Europhys. Lett. 1992, *19*, 155.
10. Español, P.; Warren, P. Europhys. Lett. 1995, *30*, 191.
11. Kong, Y.; Manke, C.W.; Madden, W.G.; Schlijper, A.G. J. Chem. Phys. 1997, *107*, 592.
12. Schlijper, A.G.; Hoogerbrugge, P.J.; Manke, C.W. J. Rheol. 1995, *39*, 567.
13. Shillcock, J.C.; Lipowsky, R. J. Chem. Phys. 2002, *117*, 5048.
14. Den Otter, W.K.; Clarke, J.H.R. J. Mod. Phys. 2000, *II*, 1179.
15. Español, P. Phys. Rev. E. 1995, *52*, 1734.
16. Marsh, C.A.; Backx, G.; Ernst, M.H. Europhys. Lett. 1997, *38*, 411.
17. Flekkøy, E.G.; Coveney, P.V. Phys. Rev. Lett. 1999, *83*, 1775.
18. Flekkøy, E.G.; Coveney, P.V.; De Fabritiis, G. Phys. Rev. E. 2000, *62*, 2140.
19. Serrano, M.; De Fabritiis, G.; Español, P.; Flekkøy, E.G.; Coveney, P.V. J. Phys. A: Math. Gen. 2002, *35*, 1605.
20. Coveney, P.V.; Español, P. J. Phys. A: Math. Gen. 1997, *30*, 779.
21. Lowe, C.P. Europhys. Lett. 1999, *47*, 145.
22. Español, P. *Simulate Challenges in Molecular Simulations*; 2002, *4*, 59; Chap. III.
23. Groot, R.D.; Warren, P.B. J. Chem. Phys. 1997, *107*, 4423.
24. Gmela, H.; Öttinger, H.C. Phys. Rev. E. 1997, *56*, 6620.
25. Öttinger, H.C.; Grmela, H. Phys. Rev. E. 1997, *56*, 6633.
26. Öttinger, H.C. Phys. Rev. E. 1998, *57*, 1416.
27. Pagonabarraga, I.; Hagen, M.H.J.; Frenkel, D. Europhys. Lett. 1998, *42*, 377.
28. Marsh, C.A.; Backx, G.; Ernst, M.H. Phys. Rev. E. 1997, *56*, 1676.
29. Español, P.; Revenga, M. Phys. Rev. E. 2003, *67*, 26,705.
30. Ripoll, M.; Ernst, M.H.; Español, P. J. Chem. Phys. 2001, *115*, 7271.
31. Evans, G.T. J. Chem. Phys. 1999, *110*, 1338.
32. Ripoll, M.; Español, P.; Ernst, M.H. Int. J. Mod. Phys. C. 1998, *9*, 1327.
33. Español, P. Europhys. Lett. 1997, *39*, 606.
34. Bonet-Avalós, J.; MacKie, A.D. Europhys. Lett. 1997, *40*, 141.
35. Bonet-Avalós, J.; MacKie, A.D. J. Chem. Phys. 1999, *11*, 5267.
36. Warren, P. Curr. Opin. Colloid Interf. Sci. 1998, *3*, 620.
37. Revenga, M.; Zúñiga, I.; Español, P. Computer Phys. Commun. 1999, *121–122*, 309.
38. Serrano, M.; Español, P.; Zúñiga, I. Computer Phys. Commun. 1999, *121–122*, 306.
39. Español, P.; Serrano, M. Phys. Rev. E. 1999, *59*, 6340.
40. Pagonabarraga, I.; Frenkel, D. J. Chem. Phys. 2001, *115*, 5015.
41. Schlijper, A.G.; Manke, C.W.; Madden, W.D.; Kong, Y. Int. J. Mod. Phys. C. 1997, *8*, 919.
42. Novik, K.E.; Coveney, P.V. Int. J. Mod. Phys. C. 1997, *8*, 909.
43. Novik, K.E.; Coveney, P.V. Phys. Rev. E. 2000, *61*, 435.
44. ten Bosch, B.I.M. J. Non-Newtonian Fluid Mech. 1999, *83*, 231.
45. Masters, A.J.; Warren, P.B. Europhys. Lett. 1999, *48*, 1.
46. Trofimov, S.Y.; Nies, E.L.F.; Michels, M.A.J. J. Chem. Phys. 2002, *117*, 9383.

47. Willemsen, S.M.; Vlugt, T.J.H.; Hoefsloot, H.C.J.; Smit, B. J. Comp. Phys. 1998, *147*, 507.
48. Groot, R.D. J. Chem. Phys. 2003, *118*, 11265.
49. Zhang, K.; Manke, C.W. Mol. Simulat. 2000, *25*, 157.
50. Kong, Y.; Manke, C.W.; Madden, W.D.; Schlijper, A.G. Int. J. Thermophys. 1994, *15*, 1093.
51. Schlijper, A.G.; Hoogerbrugge, P.J.; Manke, C.W. J. Rheol. 1995, *39*, 567.
52. van Vliet, R.E.; Dreischor, M.W.; Hoefsloot, H.C.J.; Iedema, P.D. Fluid Phase Equilibria 2002, *201*, 67.
53. van Vliet, R.E.; Hoefsloot, H.C.J.; Iedema, P.D. Polymer 2003, *44*, 1757.
54. Malfreyt, P.; Tildesley, D.J. Langmuir 2000, *16*, 4732.
55. Irfachsyad, D.; Tildesley, D.; Malfreyt, P. Phys. Chem. Chem. Phys. 2002, *4*, 3008.
56. Spenley, N.A. Europhys. Lett. 2000, *49*, 534.
57. Gibson, J.B.; Chen, K.; Chynoweth, S. J. Colloid. Interf. Sci. 1998, *206*, 464.
58. Elliott, J.A.; Windle, A.H. J. Chem. Phys. 2000, *113*, 10367.
59. Whittle, M.; Dickinson, E. J. Coll. Interf. Sci. 2001, *242*, 106.
60. Yamamoto, S.; Maruyama, Y.; Hyodo, S. J. Chem. Phys. 2002, *116*.
61. Venturoly, M.; Smit, B. Phys. Chem. Commun. 1999, *10*.
62. Kranenburg, M.; Venturoli, M.; Smit, B. Phys. Rev. E. 2003, *67*60, 901(R).
63. Jury, S.; Bladon, P.; Cates, M.; Krishna, S.; Hagen, M.; Ruddock, N.; Warren, P. Phys. Chem. Chem. Phys. 1999, *1*, 2051.
64. Bausch, A.R.; Möller, W.; Sackmann, E. Biophys. J. 1999, *76*, 573.
65. Bausch, A.R.; Ziemann, F.; Boulbitch, A.; Jacobson, K.; Sackmann, E. Biophys. J. 1998, *75*, 2038.
66. Schmidt, F.G.; Ziemann, F.; Sackmann, E. Eur. Biophys. J. Lett. 1996, *24*, 348.
67. Ziemann, F.; Rädler, J.; Sackmann, E. Biophys. J. 1994, *66*, 2210.

11

Crystallization of Bulk Fats Under Shear

GIANFRANCO MAZZANTI and ALEJANDRO G. MARANGONI
University of Guelph, Guelph, Ontario, Canada

SARAH E. GUTHRIE and STEFAN H. J. IDZIAK University of Waterloo, Waterloo, Ontario, Canada

ERIC B. SIROTA ExxonMobil Research and Engineering Company, Annandale, New Jersey, U.S.A.

I. INTRODUCTION

A vast empirical lore on the effect of shear, often secretive, exists in manufacturing facilities where bulk fats are processed. Only recently, a deeper scientific effort has been undertaken to elucidate the principles underlying these effects. In this chapter, we present the state of the art in scientific findings on the effects of shear on crystallization of fats from the melt.

The sensorial attributes of many common foods such as chocolate, butter, margarine, and spreads are derived from the structure created by their network of crystalline fat. Shear is used in regular industrial manufacturing practice because most fat systems are crystallized while being mixed to enhance the rate of heat transfer and to obtain a homogeneous final product. That is the case of chocolate, where proper mixing and tempering of the melt is essential for the manufacture of a good quality confection [1]. Similarly, processing conditions profoundly affect the consistency and spreadability of butter and margarine. The widely employed two-step process uses a scraped-surface cooler for crystallization and a mechanical finger unit to work the paste. Both machines apply high-shear conditions to the fat. Shortenings and lard products, intended to be creamed with flours and other ingredients, require an appropriate texture as well. The texture and stability of these materials are primarily influenced by the amount of solid-crystalline mass present. However, for a given solid-fat content, the internal structure of the

crystals and the microstructure of the fat crystal network can create very different mechanical characteristics [2–4]. Cooling and shear rates, the degree of undercooling, and annealing time all strongly influence the structure of the crystal network formed and thus its mechanical properties. Another important crystallization process is the fractionation of fats, especially of palm oil. Low shear rates and slow cooling rates are used to obtain an appropriate composition of crystals, whereas targeting a size distribution and mechanical strength that make the separation processes more efficient.

Triacylglycerols are the main component of edible fats. Three fatty acids are esterified on a glycerol backbone, allowing for different combinations of saturated and unsaturated hydrocarbon chains to form. The triglyceride molecules crystallize, forming one long chain that includes two of the fatty acids, and a side chain with the remaining fatty acid, in a conformation similar to an "h." Natural fats contain a wide variety of fatty acids, depending on the type of fat, its origin, and the methods used for extraction and purification. These fat systems display a rich and complex crystalline polymorphism, strongly dependent on heat, mass, and momentum transfer conditions during crystallization. Minor components, especially polar lipids, act as impurities that modify the phase behavior of the fats and are often used as natural additives to improve properties of the materials. The three main fat crystal polymorphs are, in order of increasing stability, the α (hexagonal subcell), several β' (orthorhombic subcell), and several β (triclinic subcell) forms [5,6]. These structures, and their crystallization behavior, bear strong resemblance to those of alkanes [7–9] because the long hydrocarbon chains are the dominant portion of the moluecules. The characterization of the crystalline structure of these polymorphs is done using X-ray diffraction. Although very few exact structures have been completely elucidated, the characteristic signatures of the different polymorphs have been widely documented. The ideal crystal type for chocolate is in a β_V form, because it is quite stable and melts at the right temperature in the mouth. Margarines and shortenings contain usually a mixture of β' and β crystals. The crystal size and type is very important for the structure of margarine because an excess of β crystals gives a sandy texture, and makes it difficult to spread and melt.

The effects of scraped heat exchangers and of working devices, the commonly known Votator units A and B, include mixing and homogenizing of the sample, as well as breaking of crystals and clusters. They also provide thorough mixing for proper distribution of the air (or nitrogen) and of the aqueous phase of many products. The shear patterns present in these machines are very complicated. The different regimes that affect the system arise not only from the geometry and dynamics of the device but also from the changes in the material, which enters as a liquid, becomes a suspension, and ends as a fluid paste. However, it is not immediately intuitive how this

mechanical work produces a direct effect on the phase change of solid-crystalline particles. The improved heat and mass transfer conditions can certainly account partially for the differences in phase transition kinetics observed between quiescent and sheared systems. However, the changes are often too dramatic to be explained only on the basis of those enhancements.

From the basic scientific understanding, this is a very new field and previously unknown and important phenomena have recently been shown to be common to major commercial fats. One is the orientation of crystallites formed during crystallization from the melt under the influence of shear [10]. Another is the shear-induced acceleration of monotropic phase transitions from less stable to more stable crystalline polymorphic forms, which appears also as a universal effect of shear applied to crystallizing fats [10].

II. EXPERIMENTAL PROCEDURES USED IN THE STUDY OF SHEAR EFFECTS

The methods used to observe the effects of shear on the crystallization of fats include rheology, light scatter, differential scanning calorimetry (DSC), and X-ray diffraction. Recently, the use of synchrotron radiation has permitted time-resolved X-ray studies. Many time-resolved X-ray diffraction studies of fats have been conducted under static conditions [11–18]. Studies of shear effects on the crystallization behavior and structure of food fats using a direct structural probe such as X-ray diffraction or neutron scattering are in their infancy. Only two such studies appeared in the literature [10,19], proving that cocoa butter suffered a major modification in its phase transition behavior when shear was applied [20,21] and how these effects extend to other fats [10].

In 1962, Feuge [20] showed that mechanical working of cocoa butter produced a fast change into a more stable polymorph. The same behavior was observed in other cocoa butter-like systems. Mechanical extrusion followed by X-ray diffraction of the sample was used in this study. The use of rheology and DSC to follow changes in the crystallizing mass allowed Ziegleder [21] to study the effect of shear on the induction times for the β_V polymorph in cocoa butter with more detail and under better control of the shear conditions. The dramatic reduction in times was summarized in a log–log plot of the times versus the shear rate applied, yielding a straight line for each temperature. MacMillan et al. [19] used synchrotron radiation to observe the polymorphic transformation of cocoa butter under low-shear regimes. Mazzanti et al. [10] extended the synchrotron studies to high shear and to other commercial fats such as milk fat, cocoa butter, and palm oil. Stripped milk fat was also used, having its minor polar lipids removed to determine their influence on crystallization behavior. The polar components were removed using a Florisil®

column [22]. Martini et al. [23] studied induction times for blends of milk fat fractions and hydrogenated sunflower oil mixtures using light scattering.

The time-resolved synchrotron X-ray diffraction (XRD) experiments used Couette [10] or cone-plate [19] geometries to attain a controlled shear. Each one of these geometries allows for the probing of different planes of orientation.

In the experiments done by Mazzanti et al. [10], dynamic and static experiments were carried out in a Couette shear cell (Fig. 1a). It consists of two concentric cylinders made of 0.5-mm-thick Lexan™ walls. The internal cylinder (39 mm in diameter) is stationary and has water flowing inside for temperature control. The outer cylinder (41 mm in diameter) rotates, driven by a controlled motor, to produce a uniform shear in the sample fluid placed in the 1-mm gap between the two cylinders. The path of the X-ray beam is shown in transmission through the edge of the Couette cell. Note that this

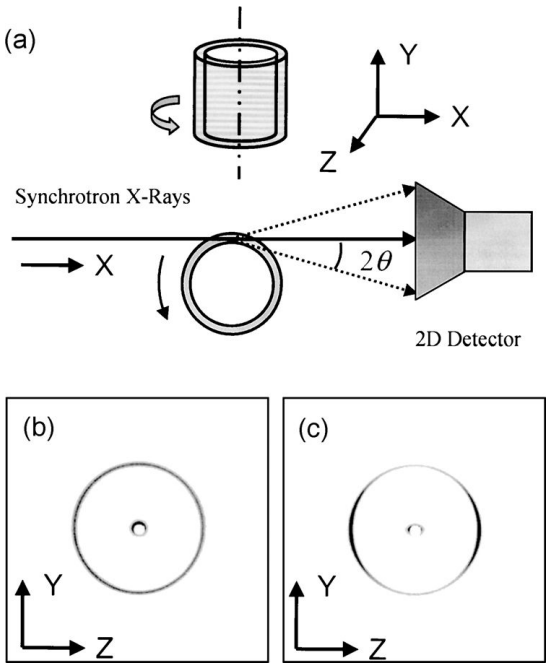


Figure 1 Couette flow cell and scattering geometry: (a) Schematic of the Couette flow cell; (b) typical two-dimensional (2D) diffraction pattern of the (001) peak of an unoriented cocoa butter in the β'_{III} phase; (c) typical 2D diffraction pattern of oriented β'_{III} cocoa butter crystals. The (001) plane normal is oriented along the velocity gradient direction, with (001) planes parallel to the walls of the shear cell.

geometry only allows for detection of structure in the X–Y and X–Z planes to be observed. Orientation perpendicular to the beam (Y–Z plane) was inaccessible, but previous work by MacMillan et al. [19] in this plane did not show orientational effects. The size of the X-ray beam was 0.5 × 0.5 mm and its wavelength was $\lambda = 1.097$ Å. A Bruker 1500 two-dimensional (2D) detector was used to capture diffraction patterns. The distance to the sample was 1129 mm for small angles. The data acquisition time was 50 s per diffraction pattern. The XRD measurements were conducted at Exxon Mobil beamline X10A at the National Synchrotron Light Source (NSLS). Typical diffraction patterns are shown in Figs. 1b and 1c. This setup allowed the observation of time-dependent crystalline growth, determination of the polymorphic form, and detection of orientation. This was accomplished primarily by examination of the (001) diffraction peak, which gives the crystal lattice spacing along the c axis of the unit cell and is characteristic of each phase. Wide-angle scattering was also used to differentiate among the α, β′ and β forms [5,6].

The fats were heated to 50°C and shear rates ranging from 45 to 2880 s^{-1} were then continuously applied to the melt. The samples were kept liquid at 50°C for 30 min under shear before the temperature was decreased at a rate of 3°C/min to the final crystallization temperature. Typical values for this temperature are 17.5°C for milk fat, stripped milk fat, and palm oil and 20°C for cocoa butter. The diffraction patterns were collected continuously at 1-min intervals.

The work done by MacMillan et al. [19] included experiments on cocoa butter crash-cooled to different temperatures and shear rates ranging from 3 to 12 s^{-1}. It also included tempering processes reheating between 30°C and 40°C.

III. CRYSTALLINE ORIENTATION

Shear-induced orientational effects have been studied extensively in complex fluids, such as liquid crystals [24–26] and polymers [27]. There has been very little research on the orientation effects of shear on fat crystals. The orientation of crystals under shear is a consequence of the asymmetry of the crystals and the intensity of the shear field. The orientation of the crystallites at the onset of the crystallization is a consequence of the shear in the liquid phase. However, as crystallization progresses, the orientation becomes also dependent on the interaction between crystallites. The microscopic observation of these crystals shows that they tend to aggregate. Therefore, the orientation is a consequence of the different forces acting upon the crystal under the different conditions imposed at different stages of the crystallization process.

A typical diffraction pattern for cocoa butter [10], crystallized statically from the melt (Fig. 1b), clearly shows the isotropic (001) diffraction peak

characteristic of unoriented material in the β'_{III} phase, one of its two possible β' forms. The peak is located at a radial position $q = 0.126$ Å^{-1} ($q = 2\pi/d$), where d is the spacing between the (001) planes. This is the distance between the terminal methyl groups of the tryacylglycerols. This pattern is also typical of the other fats, although the exact radial position depends on the chemical composition. Figure 1c shows the diffraction pattern of the same β'_{III} cocoa butter crystallized under dynamic conditions at 1440 s^{-1}. The anisotropy of the scattering intensity around the ring clearly indicates crystallite orientation. The evidence of orientation in the X-ray patterns comes primarily from the diffraction peak produced by the crystalline planes of the fat crystals. This produces an asymmetric ring of scatter. In addition to that, the orientation is also evident in the particle scatter at smaller angles. Asymmetric particles with a preferred orientation scatter X-rays in an asymmetric pattern. Observation of that scatter near the beamstop in Fig. 2 shows clearly that the particles have a preferred orientation. Therefore, because there is orientation both in the

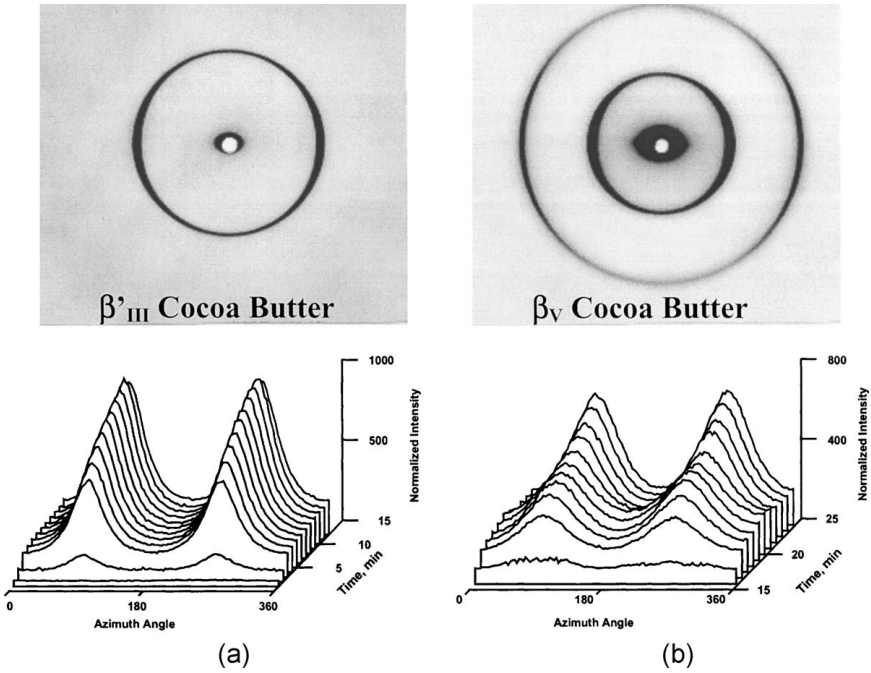


Figure 2 Two-dimensional diffraction pattern of (a) oriented β'_{III} and (b) β_V cocoa butter crystals, cooled under a shear rate of 1440 s^{-1}. The orientational azimuthal plots were taken along the (001) reflection, starting at the uppermost part of the ring.

particle scatter and in the crystalline scatter, we must conclude that the orientation is related to their internal crystalline morphology, which produces a growth mode intrinsically asymmetric in the crystallites. This effect remains apparent after the phase transition to the β_V form, which gives chocolate its ideal sensory attributes (Fig. 2b). Based on these observations, oriented chocolate was obtained following the same procedures. Figures 3a and Fig 3b show typical diffraction patterns for a commercially available dark chocolate cooled under a shear rate of 360 s^{-1}. The β'_{III} is present in Fig. 3a and the β_V form in Fig. 3b. These diffractograms show considerably more intense small-angle scatter, due to the presence of particulate matter in the chocolate coming from the cocoa powder, the milk proteins, and the sugar added in the formulation. The linear orientational plots (or azimuthal plots) derived from the 2D images correspond to changes in the normalized intensity around the ring as a function of time. As can be seen, the orientation is present since the first stages and is present throughout the growth of the material.

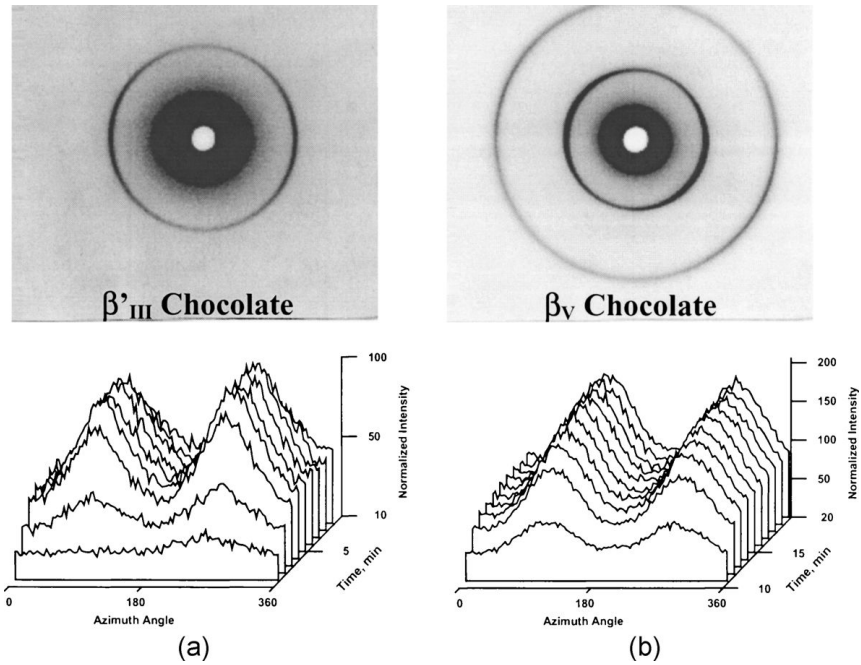


Figure 3 Two-dimensional diffraction pattern for commercial dark chocolate, cooled under a shear rate of 360 s^{-1}. The orientation of the β'_{III} and β_V crystals is apparent in (a) and (b). The tilt in the orientation is due to a secondary flow pattern developed by the chocolate paste.

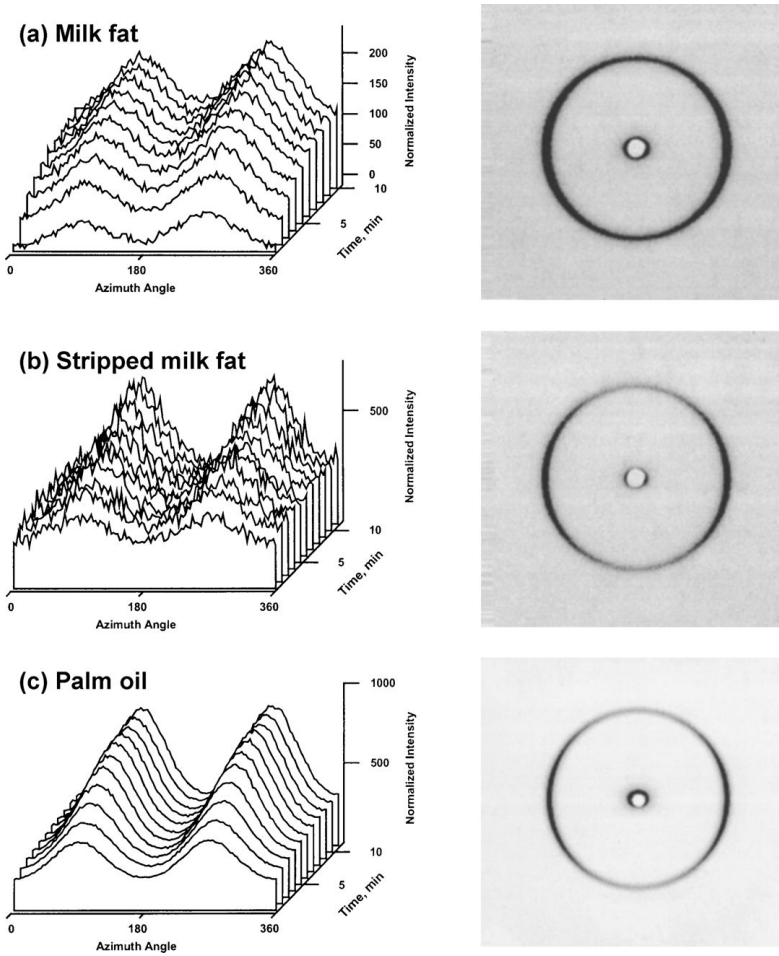


Figure 4 Time evolution of the orientational plots of the β′ poymorph upon crystallization of milk fat, stripped milk fat, and palm oil (a–c). The fats were heated to 50°C for 30 min and cooled at a rate of 3°C/min. A shear rate of 1440 s^{-1} was applied throughout the whole experiment. Milk fat, stripped milk fat, and palm oil were cooled down to 17.5°C and cocoa butter down to 20°C. The orientation of the crystals is readily apparent in all of the samples. Note that the integrated intensity increases with time as additional material is crystallized.

Additional orientational plots are shown in Fig. 4 for milk fat, milk fat stripped from its polar lipids, and palm oil for the β′ polymorph, crystallized under a shear rate of 1440 s^{-1}. The intensity in these scans is increasing with time as additional material is crystallized from the melt after the apparition of the phase. Orientational effects are evident from the onset of crystallization. Moreover, all four fat systems display similar behavior, strongly suggesting that the shear-induced effect is universal to fats. Minor polar lipids, naturally present in milk fat, reduce orientation, as shown by the narrower width of the orientation peak observed in stripped milk fat.

It can be seen from Figs. 2–4 that each material has a differet tendency to orient in the shear field. This is reasonable because the size of the crystallites and the amount of total solid material present is different for each system at a given shear rate and temperature. Cocoa butter displayed, in general, the strongest orientation, whereas palm oil was often the less oriented material. A quantitative evaluation of the orientation was done using an order parameter S, defined as the first nonzero moment of the distribution of the angles of the particles with respect to the flow direction, $S = \frac{1}{2}\langle 3\ cos^2\chi - 1\rangle$. A material with order parameter S below 0.02 is virtually not oriented. A perfectly oriented material would have an order parameter of 1. The increase in orientation with shear [28] is presented in Fig. 5. It can be seen that at high-shear rates, milk fat, stripped milk fat, and palm oil achieved approximately the same orientation. Cocoa butter started at a lower level that stripped milk fat but eventually reached a higher orientation.

The importance of the observation of the orientational effects is twofold. First, it offers a clear confirmation that the shear forces acting on the surface of separate crystallites can be significant enough to modify the interaction between crystallites and between a crystallite and its surrounding liquid. This concept may be very important when trying to understand the effect that shear has on the phase behavior of fats. Second, they show that there is a particular fundamental crystalline structure limited to small particles that tend to aggregate to form clusters, which, in turn, generate the actual structure of the fat. The structuring of bulk fats using shear, a common semiempirical practice, can then be approached with similar tools used to understand the structures of fats crystallized statically [29], but enriched now by this new understanding of the way that the basic particles of the system tend to interact when subjected to shear.

As was pointed out earlier, during crystallization the material goes through different stages, beginning with nucleation and followed by unhindered growth and structure formation. The initial nucleation is characterized by the appearance of nuclei relatively far apart from each other. These platelet-like nuclei [30] do not to affect each other. As they grow, the system becomes a disperse suspension of rapidly growing crystals. Eventually, the crystals form

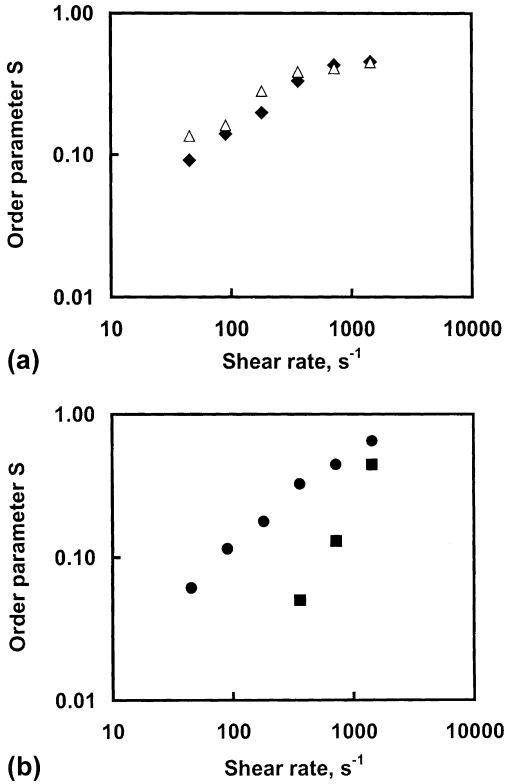


Figure 5 Effect of shear on the orientation of crystallites of the β′ polymorph of milk fat and stripped milk fat (a) and of cocoa butter and palm oil (b). ◆: Milk fat; △: stripped milk fat; ●: cocoa butter; ■: palm oil.

clusters that impinge upon each other, as shown by polarized light microscopy observations under static conditions (Fig. 6) [10]. As the small crystallites agglomerate, they squeeze out the oil in between them and the clusters become more compact and age. Note that these aggregates, crystallized under static conditions, do not show the radial distribution characteristic of spherulites (manifested by a "Maltese Cross" under polarized light), but rather a random orientation of the crystallites within the spherical cluster, as seen in Fig. 6.

Moderate shear fields seem to be able to prevent the formation of these clusters, although the exact response depends on the chemical composition of the system and processing conditions. Some results are consistent with the formation of small asymmetric crystals, which in the presence of a shear field adopt a nonrandom distribution around an average preferred orientation

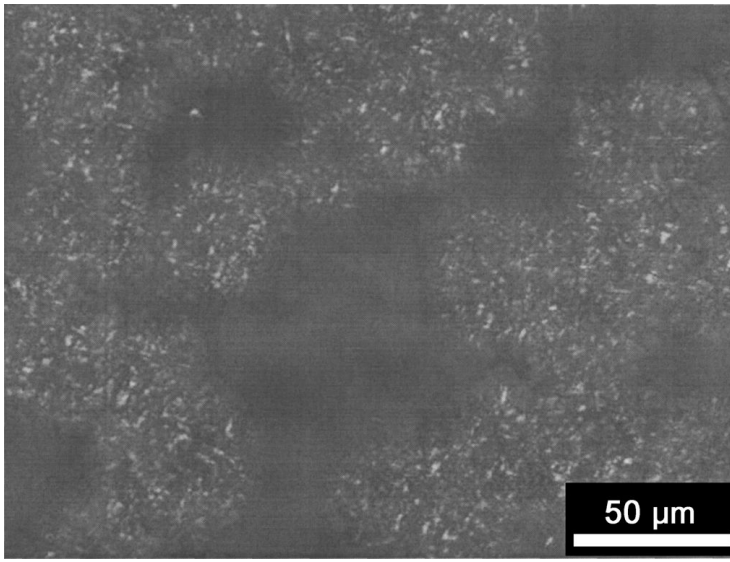


Figure 6 Polarized light micrograph of cocoa butter crystallized at 20°C after 1 day under static conditions. The clusters of size 40–50 μm are formed by agglomeration or secondary nucleation of smaller crystallites of 1–2 μm. These spherical clusters are not spherulites because they do not show the characteristic Maltese Cross radial pattern. The clusters are composed of crystallites mostly in the β'_{IV} phase.

(Fig. 7a). On the other hand, weak or no orientation was observed at low-shear rates either due to a random distribution of anisotropic crystals (Fig. 7b) or the formation of spherical particles upon platelet aggregation (Fig. 7c).

The different agglomeratin behaviors with or without shear would explain the differences observed when fats were crystallized for 1 h at 90 s^{-1}, followed by 10 min at 1440 s^{-1}. Milk fat and cocoa butter crystals readily became oriented, suggesting a transition between disordered (Fig. 7b) and oriented (Fig. 7a) states. Palm oil crystallized statically and sheared afterward, however, did not become oriented, thus implying a spherical crystalline particle structure (Fig. 7c) similar to the static situation (Fig. 6) but with the rounded clusters rolling in the shear field.

The degree of orientation of suspended particles in a flowing system depends on the interaction of shear, interparticle, and Brownian forces [31]. In low-volume-fraction particle suspensions, like the ones present at nucleation and the early stages of crystallization, interparticle forces are negligible. Therefore, the distribution of particle orientations results from the interplay between ordering induced by shear forces and disordering induced by

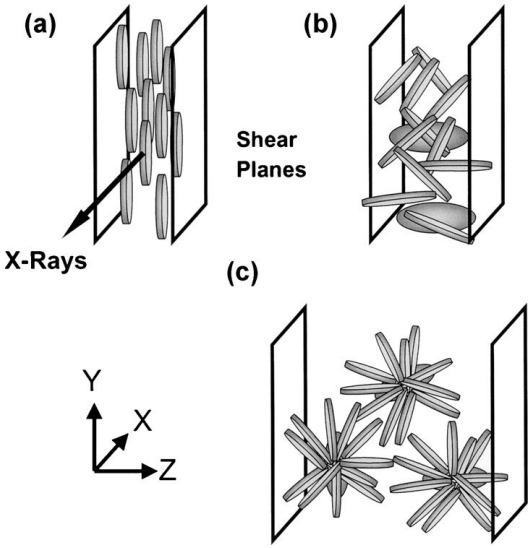


Figure 7 Possible orientations of crystals: (a) sketch of the oriented asymmetric platelike crystals formed while crystallizing under shear; (b) the crystallites lose orientation if the shear is reduced; (c) spherical particles formed at low shear. The platelike crystals aggregate and cannot be shear oriented due to their isotropic conformation.

Brownian forces, resulting from thermal motion of the molecules. If shear forces are prevalent, an asymmetric particle rotates slowly when its longer axis is nearly parallel to the direction of flow, and much faster when it is perpendicular to it. The resulting time-averaged distribution is then preferentially parallel to the direction of flow, and this is what the detector captures over a given time window. Because tryacylgliceride systems tend to crystallize, initially forming small platelets [30], it is reasonable to assume that the nuclei are plateletlike shaped and originally of only one layer of molecules across. To estimate the magnitude of the Brownian forces on such platelets, one can use the rotary diffusivity [31] $D_r = 3k_BT/\pi\eta_s d_p^3$, which is inversely proportional to viscosity (η_s) and particle diameter (d_p). The crossover between the shear regime and the Brownian regime is described by the nondimensional Peclet number (Pe) at a given shear rate $\dot{\gamma}$, $\text{Pe} \equiv \dot{\gamma}/D_r$. At values of Pe larger than 10, orientation effects become obvious for many systems [31]. The dependence of the Peclet number on particle diameter at shear rates of 90 and 1440 s^{-1} is shown in Fig. 8. The viscosity measured for the liquid fat just before nucleation was used for these computations. Depending on the size of the nuclei formed at the onset of crystallization, the particles will be under the

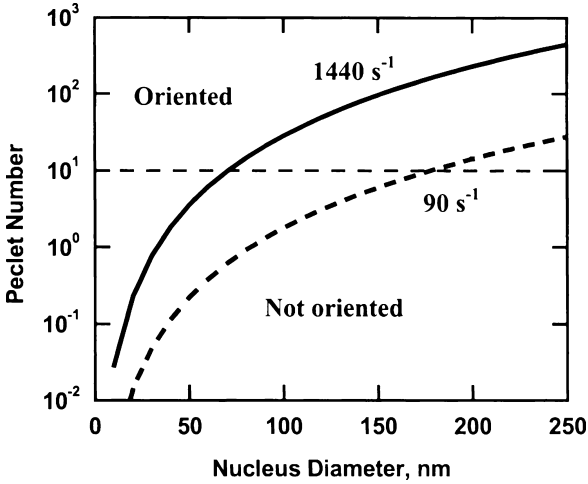


Figure 8 The measured viscosity of the oil phase just previous to the onset of crystallization ($\eta_s = 0.06$ Pa·s) was used to estimate the Peclet number as a function of shear rate and possible nucleus size, as described in the text.

shear regime or under the Brownian regime. It is possible to estimate the size of these nuclei from thermodynamic considerations and experimental data.

During nucleation, only the nuclei that reach a critical diameter d_c will grow further, and nuclei below that size will melt. An estimation of order of magnitude of the critical diameter (d_c) of a typical fat crystal nucleus is possible based on experimental data, using a modified Gibbs–Thompson approach for flat cylinders [32,33]. Assuming a monolayer platelet [30], the critical diameter can be computed as $d_c = 4\sigma_s T_f V_m / \Delta T \Delta H_f$ yielding a value of $d_c \approx 100$ nm. The values for this estimation were taken from the work of [34], in which Rousset the enthalpy of melting is $\Delta H_f = 25.6$ kJ/mol, the melting temperature $T_f = 296.5$ K, the degree of undercooling $\Delta T = 3.5$ K, the molar volume $V_m = 8 \times 10^{-4}$ m^3/mol, and the average spherical surface free energy $\sigma = 0.004$ J/m^2. The edge surface free energy was estimated as $\sigma_s = [\frac{2}{3}\sigma^3, (4T_f V_m / d\Delta H_f \Delta T)]^{1/2}$, with $\sigma_s = 0.0099$ J/m^2. In experiments on cocoa butter, a (001) interplanar spacing of $d \approx 50$ Å (0.5 nm) has been measured, which was used as the incipient crystal nucleus monolayer thickness. At the shear rate of 1440 s^{-1} studied, the minimum platelet size that will orient, with a Peclet number larger than 10, is about 70 nm (Fig. 8). The critical diameter for platelet formation was computed to be 100 nm. This suggests that the particles formed at high shear are most likely under the shear regime since the onset of nucleation, and the shear forces are able to prevent them from forming

unoriented clusters. Particles would have to grow to about 200 nm (Fig. 8) before they would be oriented by a low shear. However, if by this stage the crystallites already formed spherical clusters, they would no longer be able to be oriented, which is consistent with the observed weak orientation at low-shear rates in some experiments. Therefore, when the templating process takes place in the vicinity of the nucleation event, it is possible to form a structure made primarily of many small crystallites, rather than made up of clusters of crystallites. A later increase of shear may disrupt the aggregates formed at low or no shear, allowing the separated crystallites to reorient again. The use of shear to break clusters of crystals into smaller units is quite common in many crystallization processes.

The shear forces that induce orientation are applied to the crystal surface and they may be responsible for the acceleration of the phase transitions. It can be speculated that these forces either induce a solid-state transformation, where various polymorphs are very close in free energy (Figs. 9a–9d) or melt most of the crystal, leaving behind a nucleus for the growth of a more stable phase, as seems to be the case when a remainder of the unstable form persists for a longer time. However, the acceleration of phase transitions in cocoa butter has been shown to occur at shear rates as low as $3\ s^{-1}$ [19], and from the general observations, it is unlikely that at this shear level, the material was very oriented [10]. Nevertheless, the shear was enough to produce a dramatic change in phase behavior, most likely reducing the degree of aggregation of the crystallites into clusters. Thus, the fundamental mechanism of the phase acceleration phenomena presented in the next section remains still somewhat undefined.

IV. PHASE BEHAVIOR UNDER SHEAR

A now better known effect of shear forces at the surface of the crystals is the acceleration of the transformation to more stable polymorphs. It has been confirmed by MacMilan et al. [19] for cocoa butter, and by Mazzanti et al [10] for milk fat, stripped milk fat, cocoa butter, and palm oil, using time-resolved X-ray observations. For example, the α to β' transition times in milk fat, stripped milk fat, and palm oil were reduced by an order of magnitude in sheared samples relative to statically crystallized samples (Fig. 9). The data presented use time zero as the moment when the crystallization temperature was reached. Stripped milk fat crystallized faster than regular milk fat, thus suggesting that minor polar lipids, acting as impurities, hinder crystallization events, stabilize the α phase, and delay the formation of the more stable β' phase.

The polymorphic transformations of cocoa butter are more complex because the β'_{IV} phase in not always present, as illustrated in Fig. 9c. Under static conditions, the β'_{III} phase persists for approximately 75 min, whereas

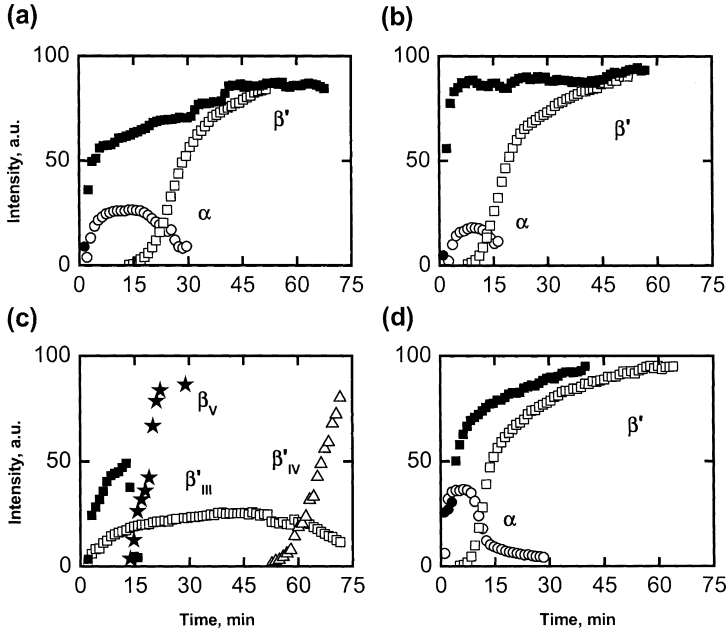


Figure 9 Crystalline content represented by the integrated X-ray peak intensity. The α (circles) and β′ (squares) are from (a) milk fat, (b) stripped milk fat, (c) cocoa butter, and (d) palm oil. Integrated peak intensity of the β'_{III} (squares), β'_{IV} (triangles) and β_V (stars) peaks for cocoa butter. Samples were melted at 50°C for 30 min and then cooled to 17.5°C (milk fat, stripped milk fat, and palm oil) or 20°C (cocoa butter) at a rate of 3°C/min. Results from static experiments correspond to the open symbols and results from the dynamic experiments correspond to the solid symbols. The accelerated transition between the phases can be clearly appreciated under shear rate of 1440 s^{-1}.

the β'_{IV} phase starts to emerge after 50 min (the α phase was not observed). The desirable β_V form of cocoa butter, present in high-quality chocolate, requires at least an additional 15 h to form through a tempering process [13,18]. However, when cocoa butter was crystallized under a shear rate of 1440 s^{-1}, the formation of the β'_{IV} phase was only observed in some experiments for a short time, whereas the β_V phase formed in less than 20 min, in agreement with previous reports [1,19]. A similar behavior was observed in chocolate, sheared at 360 s^{-1}. The persistence time of the β'_{III} phase after the beginning of the phase transition was reduced as shear was increased, and at 1440 s^{-1}, no evidence of the presence of the β'_{III} phase was observed after the phase transition. This is especially important for the confectionery industry, in which a reduction in the time required to induce the formation of desirable crystalline structures is extremely valuable.

Depending on the conditions, however, the β'_{IV} phase may shortly appear before the transition to the β_V phase happens. The permanence of the β'_{III} or β'_{IV} phase coexisting with the β_V phase may be due to a solid-state transformation.

There seems to be a general trend of reduction in phase transition times as a logarithmic or linear function of the shear rate applied for the systems documented so far [10,19,21]. However, for products that have a high-solids content, the increase in viscosity may eventually give way to an increase in the dissipated heat. This will produce local or global heating and therefore lead to a different kinetic behavior. It is clear that there is still a wide area open for scientific exploration to determine the interaction of the many mechanisms involved.

The effect of shear rate on the onset of the β' phase for milk fat, stripped milk fat, and palm oil [28] is shown in Fig. 10a. It is clear the onset time is reduced following more or less an exponential growth in the shear rate. The transition times also became shorter as the shear rate was increased. Thus, this behavior indeed seems to be common to fats of very different composition. The transitions in palm oil were noted to be smoother, even between quiescent and slow sheared conditions. Milk fat and stripped milk fat showed initially a dramatic acceleration and then followed the smooth trend presented in Fig. 10a.

Cocoa butter has the peculiarity of reaching a large solid-fat content faster; therefore, there is a notorious increase in the apparent viscosity of the paste as crystallization progresses. Therefore, the reduction in the onset times for the phase transition were observed only up to a shear rate of 360 s^{-1}, as shown in Fig. 10b. At higher shear rates, the times began to grow again, probably due to viscous heating.

Although the control of phase transitions is essential in manufacture, the microstructure is not only a consequence of the SFC and the polymorphic state of the crystal. The mixing and shearing procedures introduce a particular way of shaping the microstructure, by forming clusters of smaller size, by breaking previously formed clusters, by redistributing material that can act as a secondary nucleation agent, and by defining preferred orientations for the crystallites. These structures, that are not formed only by the quiescent redistribution of matter, will have their own set of mechanical and functional properties, which are still now obtained by laborious trial- and-error procedures combined with educated guesses.

The recent findings and the methods developed bring the task of equipment optimization and design, and the blending art common to all fat-based processors, to a stage more promising than ever. The particular behavior of each fat system still requires individual characterization, but the development of appropriate tools and understanding make it now feasible.

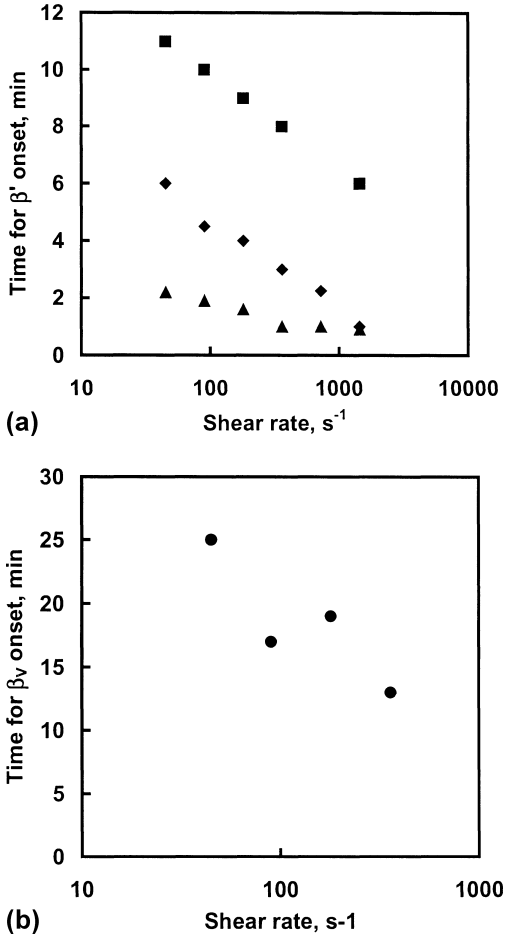


Figure 10 (a) Time for onset of the β′ polymorph in milk fat, stripped milk fat, and palm oil. (b) Time for the onset of the β_V polymorph in cocoa butter. ▲: Stripped milk fat; ◆: milk fat; ■: palm oil; ●: cocoa butter.

ACKNOWLEDGMENTS

We thank Steve Bennett for technical assistance. We acknowledge the financial assistance of the Natural Sciences and Engineering Research Council of Canada (NSERC), the Ontario Ministry of Agriculture and Food, the Ontario Ministry of Energy, Science and Technology, the Canada Research Chair Program and the Dairy Farmers of Ontario. Research carried out in part at the NSLS, BNL, which is supported by the U.S.

Department of Energy, Division of Materials Sciences and Division of Chemical Sciences. We thank Sarah McGauley for the micrograph and Kisun S. Park for computer programming.

REFERENCES

1. Beckett, S.T. *The Science of Chocolate*; Royal Society of Chemistry: Cambridge, 2000.
2. Narine, S.S.; Marangoni, A.G. Phys. Rev. E. 1999, *60*, 6991–7000.
3. Narine, S.S.; Marangoni, A.G. Phys. Rev. E. 1999, *59*, 1908–1920.
4. Marangoni, A.G. Phys. Rev. B. 2000, *62*, 13,951–13,955.
5. Sato, K.; Ueno, S.; Yano, J. Prog. Lipid Res. 1999, *38*, 91–116.
6. Sato, K. Chem. Eng. Sci. 2001, *56*, 2255–2265.
7. Sirota, E.B.; Herhold, A.B. Science 1999, *283*, 529–532.
8. Sirota, E.B. J. Chem. Phys. 2000, *112*, 492.
9. Small, D.M. *The Physical Chemistry of Lipids*: *From Alkanes to Phospholipids*; Plenum Press: New York, 1986.
10. Mazzanti, G.; Guthrie, S.E.; Sirota, E.B.; Idziak, S.H.J.; Marangoni, A.G. Crystal Growth Des. 2003, *3*, 721–725.
11. van Malssen, K.F.; Peschar, R.; Schenk, H. J. Am. Oil Chem. Soc. 1996, *73*, 1209–1251.
12. van Malssen, K.F.; Peschar, R.; Brito, C.; Schenk, H. J. Am. Oil Chem. Soc. 1996, *73*, 1225–1230.
13. van Malssen, K.F.; van Langevelde, R.; Peschar, R.; Schenk, H. J. Am. Oil Chem. Soc. 1999, *76*, 669–676.
14. van Langevelde, A.; van Malssen, K.F.; Peschar, R.; Schenk, H. J. Am. Oil Chem. Soc. 2001, *78*, 919–925.
15. van Langevelde, A.; Driessen, R.; Molleman, W.; Peschar, R.; Schenk, H. J. Am. Oil Chem. Soc. 2001, *78*, 911–918.
16. Lopez, C.; Lavigne, F.; Lesieur, P.; Bourgaux, C.; Ollivon, M. J. Dairy Sci. 2001, *84*, 756–766.
17. Lopez, C.; Lavigne, F.; Lesieur, P.; Keller, G.; Ollivon, M. J. Dairy Sci. 2001, *84*, 2402–2412.
18. Loisel, C.; Keller, G.; Lecq, G.; Bourgaux, C.; Ollivon, M. J. Am. Oil Chem. Soc. 1998, *75*, 425–439.
19. MacMillan, S.D., et al. Crystal Growth Desi. 2002, *2*, 221–226.
20. Feuge, R.O.; Landmann, W.; Mitcham, D.; Lovegren, N.V. J. Am. Oil. Chem. Soc. 1962, *39*, 310–313.
21. Ziegleder, G. Int. Z. Lebensm. Tech. Verfahrenst 1985, *36*, 412–418.
22. Wright, A.J.; Hartel, R.W.; Narine, S.S.; Marangoni, A.G. J. Am. Oil Chem. Soc. 2000, *77*, 463–475.
23. Martini, S.; Herrera, M.L.; Hartel, R.W. J. Am. Oil. Chem. Soc. 2002, *79*, 1063–1068.
24. Idziak, S.H.J., et al. Science 1994, *264*, 1915–1918.

25. Idziak, S.H.J.; Koltover, I.; Israelachvili, J.N.; Safinya, C.R. Phys. Rev. Lett. 1996, *76*, 1477–1480.
26. Safinya, C.R., et al. Science 1993, *261*, 588–591.
27. Koppi, K.A.; Tirrell, M.; Bates, F.S. Phys. Rev. Lett. 1993, *70*, 1449–1452.
28. Mazzanti, G.; Welch, S.E.; Sirota, E.B.; Idziak, S.H.J.; Marangoni, A.G. 2004. *in preparation*.
29. Marangoni, A.G. Trends Food. Sci. Technol. 2002, *13*, 37–47.
30. Unruh, T.; Westesen, K.; Bösecke, P.; Lindner, P.; Koch, M.H.J. Langmuir 2002, *18*, 1796–1800.
31. Larson, R.G. *The Structure and Rheology of Complex Fluids 263-322*; Oxford University Press: Oxford, 1999.
32. Uhlmann, D.R.; Kritchevsky, G.; Straff, R.; Scherer, G. J. Chem. Phys. 1975, *62*, 4896–4903.
33. Kraack, H.; Deutsch, M.; Sirota, E.B. Macromolecules 2000, *33*, 6174–6184.
34. Rousset, P. In *Physical Properties of Lipids*; Marangoni, A.G., Narine, S.S., Eds. Marcel Dekker: New York, 2002; 1–36.

12

Foods at Subzero Temperatures

H. DOUGLAS GOFF University of Guelph, Guelph, Ontario, Canada

I. INTRODUCTION

Food freezing has become a well-established food preservation technique. It is quite a new process, in contrast to the traditional presentation techniques, that owes its beginnings to the development of mechanical refrigeration and has really only been developed over the last half-century [1]. That it has gained such widespread importance is due to the fact that it delivers to the consumer many desirable attributes: high-quality products, diversity in meal preparation, and convenience. This latter attribute received a tremendous boost in importance with the development and widespread use of home microwave ovens, which are now seen to be an integral part of the development and home utilization of frozen foods [1].

Frozen foods have a much longer shelf life than their fresh counterparts. Preservation is achieved by two factors: lower temperature and reduction in water content [2]. It is generally recognized that detrimental chemical (e.g., oxidative), biochemical (e.g., enzymatic), or microbiological changes in foods are reduced with lowering temperature. Therefore, storage at temperature well below 0°C markedly reduces the rate at which changes occur. It is also generally recognized that water is required for most of the above changes to occur. Thus, a reduction in water content (e.g., drying) also reduces rates of reactions. During freezing, water is removed from foods by converting it to ice. Hence, freezing induces preservation for both of these reasons. However, despite the low temperatures and tremendous reduction in unfrozen water content, frozen foods possess a finite shelf life [3]. Frozen foods stored at or near the conventional −18°C storage temperature are not completely frozen, nor are they inert. Detrimental changes may have been induced by cooling or by freezing. The ice phase may have caused structural damage and is itself not stable, undergoing changes in size, shape, and location of ice crystals with temperature fluctuation and/or time [4]. Reduction of water causes freeze

concentration of components, which itself can increase reaction rates and may also bring components together through structural rearrangements that then can react or interact with negative consequences [5].

The objective of this chapter is not to provide a comprehensive review of food freezing or frozen food quality. Many good sources of information are available providing comprehensive reviews of these subjects [1,3,6]. Rather, in keeping with the general objectives of this book, this chapter will focus on structure and structural changes (dynamics) of frozen foods. The theme of this book is soft materials. Food materials, mostly being viscoelastic and having a high water content, are, by their nature, soft. Freezing converts water to ice [2], and ice, by itself, is a very hard elastic and brittle solid material [7]. It should be pointed out, though, that during freezing, only the water is converted to ice, and the other food components do not necessarily change [5]. This ice is usually in the form of discrete crystals that are embedded in the food structure. Depending on the solute concentration, there may be a very large unfrozen phase at subzero temperatures typical of frozen food storage and distribution [8]. Thus, it is still very easy to categorize a frozen food as a soft material. At some sufficiently low temperature dictated by structure and composition, the unfrozen phase can undergo a glass transition, thereby becoming an amorphous solid material [9]. In the case of a frozen food below its glass transition temperature, we have a crystalline solid—ice—embedded in an amorphous solid—glass—and at this stage we no longer truly have a soft material. This chapter will review the formation and structure of the crystalline solid phase, the resulting structure of the unfrozen phase, and its glass transition. Although a general approach will be taken, food structure varies tremendously, so structural heterogeneity between tissue-based and nontissue-based systems will also be addressed. Structure dynamics will be addressed by focusing on recrystallization of the ice phase, moisture migration, and structural rearrangements within the unfrozen phase.

II. STRUCTURE OF FROZEN SYSTEMS

Many types of food are successfully frozen and these represent a wide diversity of structures. However, the act of freezing causes commonality to exist in structure across such a diverse group. First, there is the formation of an ice phase as water crystallizes. There are two aspects of ice formation to consider. The first of these is the establishment of ice crystal size, as dictated by the processes of nucleation and growth of crystals [7,8]. The second of these is the establishment of the ice-phase volume, which is a function of both temperature and the nature of dissolved solutes present, because these solutes depress the freezing point and result in freeze concentration [8]. A second common feature of the structure of frozen foods is that the combined effect of

low temperature and freeze concentration of the unfrozen phase means that this phase can undergo a glass transition due to its very high viscosity [9]. These elements are common to frozen food structure, but the initial structures of foods before freezing varies greatly, and the effect of this initial structure on frozen food structure will also be examined in plant and animal tissues and solutions, macromolecular dispersions, and gels.

A. Formation of the Ice Phase

1. Nucleation, Growth, and Ice Crystal Size Distributions

Crystallization is a two-step process: formation of a stable crystal embryo/nucleus (nucleation) followed by growth of that nucleus by the subsequent deposition of further molecules [2,10,11]. The objective of a food-freezing process is to promote nucleation to form as many small crystals as possible [12,13]. Small crystals are associated with minimal cell and tissue disruption, minimal water migration, minimal impact on texture when foods are eaten frozen (e.g., ice cream), and, hence, longer shelf life (see Section III.A) [14]. In foods, nucleation is always heterogeneous, except for instances where secondary nucleation is used to promote very small ice crystal sizes, as in scraped-surface freezers [11,12,15]. Ice crystal size distributions are normally determined by microscopy and frequently used parameters to characterize the size distribution are mean or median diameters, often weighted based on surface area or volume rather than number distributions [16].

The formation of small stable nuclei is a sequence of bimolecular processes whereby molecules in the liquid phase join a growing cluster (embryo), which has formed as temperature is reduced and molecular motion slows down. For a cluster of radius r, the process is governed by ΔG_{ls}, the net free energy of formation accompanying the liquid-solid condensation, and is given by

$$\Delta G_{ls} = \frac{4}{3}(\pi r^3 \Delta G_v) + 4\pi r^2 \gamma \tag{1}$$

where r is the radius of the particle, ΔG_v is the difference in free energy between the solid and aqueous phases, and γ is the interfacial free energy per unit area between the ice and unfrozen phases [7,8,11,12]. In other words, the total ΔG_{ls} is the summation of surface and bulk or volume energy terms.

At temperatures below the initial melting point (undercooling), clusters of molecules have a thermodynamically unfavorable interface with the undercooled liquid. Because the surface-to-volume ratio is large at small cluster size, the total interfacial surface energy constitutes a barrier to growth. However, as the cluster size (embryo) increases, the surface energy becomes increasingly larger, whereas the volume energy term from Eq. (1) progressively decreases at a lower rate, which favors nucleation. The relationship

among the various quantities in Eq. (1) are shown in Fig. 1, demonstrating the change in ΔG_{ls} with embryo or nucleus radius for water at $-40\,^{\circ}C$. As depicted in Fig. 1, ΔG_{ls} passes through a maximum (ΔG_{ls}^{*}) at which a radius of critical size (r^{*}) exists having an equal probability of growing or disintegrating [12]. The addition of more molecules into the lattice of the cluster at r^{*} results in a reduction in ΔG_{ls} (Fig. 1) as the newly formed nuclei gain more bulk energy [first term in Eq. (1)] than that lost in surface energy. Beyond r^{*}, the cluster is considered an active nucleus and the process becomes spontaneous, leading to crystal growth. Factors affecting the size of the critical nucleus capable of promoting condensation of molecules from the liquid phase include the interfacial free energy γ, the latent heat of fusion ΔH_f, and the degree of undercooling $\Delta T = T_f - T$, where T_f is the equilibrium solid–liquid freezing temperature and T is the system temperature [7,8,11,12]. These variables are related as follows:

$$r^{*} = \frac{2\gamma T_f}{\Delta H_f, \Delta T} \tag{2}$$

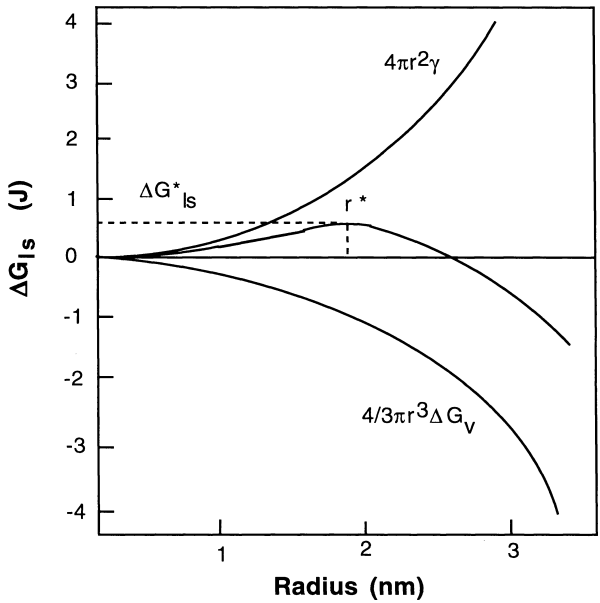


Figure 1 Changes in volume free energy of nucleation (ΔG_{ls}) as a function of cluster radius (r) for pure water at $-40\,^{\circ}C$. The surface ($4\pi r^2\gamma$) and volume $\left[\frac{4}{3}(\pi r^3 \Delta G_v)\right]$ energy terms are additive leading to the maximum seen in ΔG_{ls} at the critical radius, r^{*}.

The relationship in Eq. (2) relies on several assumptions, two of which are that the thermodynamics at the macroscopic level applies to microscopic systems and that γ is independent of r and exhibits properties similar to an ice–water interface at 0°C. Although these assumptions create uncertainty in the absolute value of r^*, it has been shown that r^* is dependent on ΔT (Fig. 2) [7,8,11,12]. As undercooling or the rate of heat removal increases (greater ΔT), smaller nuclei are stable and can exist in equilibrium with the undercooled water. It also follows that with increased undercooling, the number of water molecules at r^* is reduced.

In contrast to nucleation, the rate of crystal growth decreases with increased rate of heat removal or increased undercooling (Fig. 3) [2,7]. This is the molecular explanation for the empirical observation that faster freezing rates give rise to smaller, more numerous crystals. It also explains the industrial practice of trying to maximize the freezing rate by employing low temperatures and maximizing ΔT, exposed surface areas of food, and convective heat transfer coefficients through high air velocity [2,13,17,18].

2. Effect of Solutes, Freeze Concentration, and Ice-Phase Volume

Although ice crystal size distributions are important to frozen food quality, so too is the total ice-phase volume formed as a function of temperature. The presence of solutes in food systems results in depression of the freezing point based on Raoult's law, which relates vapor pressure of the solution to that of pure solvent based on solute concentration (Fig. 4) [2,7,8,12]. During the

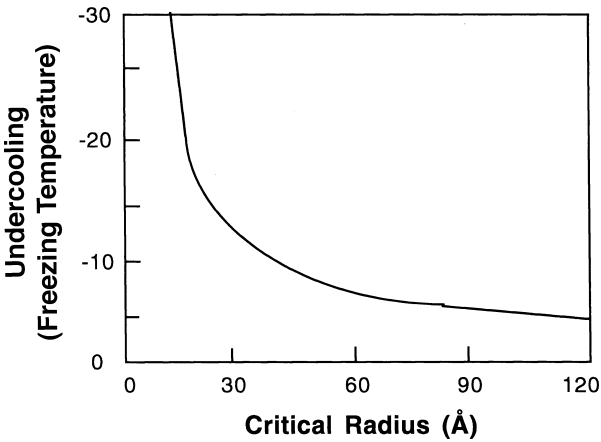


Figure 2 Change in critical radius for nucleation of water as a function of undercooling temperature.

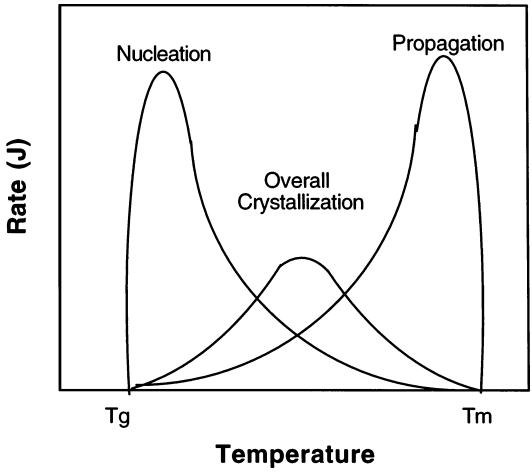


Figure 3 Comparison of nucleation and propagation rate profiles as a function of temperature. Curves are additive, leading to an overall crystallization rate.

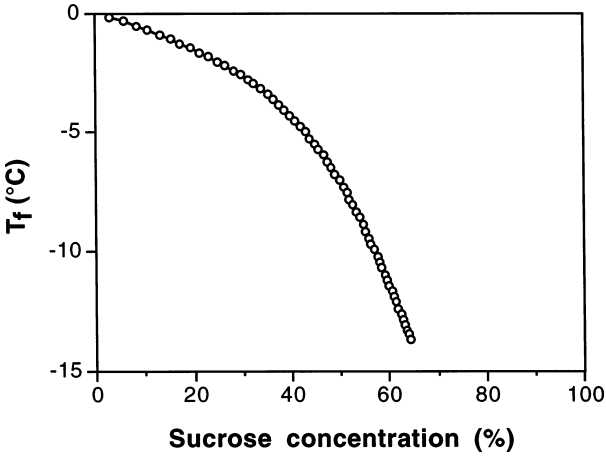


Figure 4 Freezing temperature (T_f) of sucrose solutions as a function of concentration (data from Ref. 19). Freeze concentration also proceeds along the same line from any given starting concentration.

freezing of aqueous solutions, a freeze-concentration process occurs as water freezes out of the solution in the form of pure ice crystals [11]. This causes the freezing temperature of the remaining solution to drop. Continued crystallization of the solvent can only occur as temperature is decreased (Fig. 4). At temperatures well below the initial freezing point, some liquid water remains. Enthalpy–temperature relationships for water and an aqueous solution are shown in Fig. 5 [2]. As the solution concentration increases, equilibrium T_f and ΔH_f decrease. The ice-phase volume as a function of temperature in aqueous solutions can be determined based on the enthalpy difference between the semifrozen solution and the fictitious enthalpy of the unfrozen solution at the same temperature, as determined by extrapolation [20]. Also, with increased concentration and decreased temperature, a large increase in the viscosity of the unfrozen phase occurs, thus decreasing the diffusion properties of the system and hindering crystallization [21,22]. Complete crystallization of the solute never occurs, as low temperature and high viscosity lead to a glass transition in the freeze-concentrated unfrozen phase (UFP).

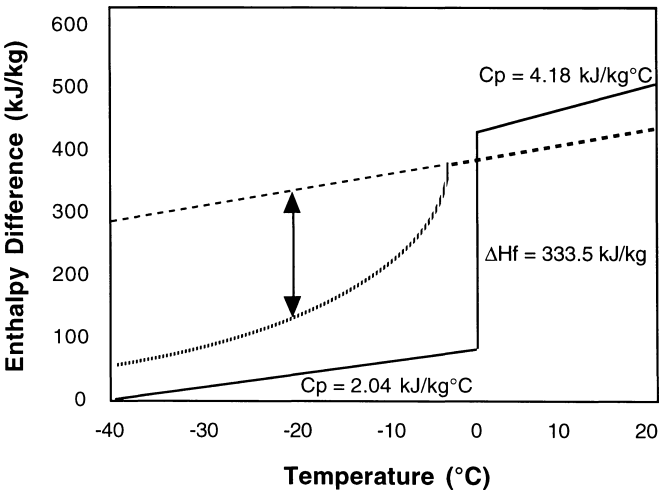


Figure 5 Enthalpy as a function of temperature for pure water (solid line), showing the heat capacities (C_p) of water and ice and the latent heat of freezing (ΔH_f), and for an aqueous solution (bold dashed line). If the enthalpy line above T_f for the aqueous solution is extrapolated to $T < T_f$ (thin dashed line), the difference in enthalpy between that and the actual enthalpy content at any T (e.g., arrowed line) is an indirect measure of the ice content at that T. (From Ref. 20.)

B. Glass Transition of the Unfrozen Phase

The glass transition associated with the freeze-concentrated UFP in frozen foods has received considerable research attention in the last several years [23–35]. Quality losses (e.g., enzymatic reactions, recrystallization) in frozen foods may be greatly reduced or inhibited when the UFP is in the glassy state [18]. Knowledge of the influence of environmental and indigenous factors (viz. temperature, freezing rate, composition) on the glass transition may make possible the formulation, processing, or storage of frozen foods in a manner that will enhance shelf life [18]. As temperature is lowered beyond solute maximum solubility concentrations (eutectic concentrations), freeze concentration proceeds in a nonequilibrium state [9]. It is unlikely, however, that solutes will crystallize due to high viscosity and low temperature. The highly concentrated UFP can then go through a viscous liquid–glass-state transition, driven by the reduction in molecular motion and diffusion kinetics [36,37]. A glass is defined as a nonequilibrium, metastable, amorphous, disordered solid of extremely high viscosity (i.e., 10^{10}–10^{14} Pa·s), also a function of temperature and concentration [9,11,12]. The glass transition curve extends from the glass transition temperature (T_g) of pure water (−134°C) to the T_g of pure solute. The equilibrium-phase diagram and the glass transition can be modeled together on a supplemented state diagram [9,38–40]. The supplementary state diagram showing the solid–liquid coexistence boundaries and glass transition profile for a binary sucrose–water system is shown in Fig. 6 [12]. Below and to the right of the glass transition line, the solution is in the amorphous glass state, with or without ice present, depending on temperature and freezing path followed, whereas above and to the left of the glass transition line, the solution is in the liquid state, with or without ice, depending on temperature.

As an example, assume a sucrose solution with an initial concentration of 20% at room temperature (point A of Fig. 6). The initial T_g of this solution at room temperature before phase separation is marked as point B (if the solution could be undercooled to this temperature without ice formation). However, upon slowly cooling of the system to somewhat below its equilibrium freezing point (point C, due to undercooling), nucleation and subsequent crystallization begins at slightly below point D and initiates the freeze-concentration process, removing water in its pure form as ice. As ice crystallization proceeds, the continual increase in solute concentration (removal of water) further depresses the equilibrium freezing point of the UFP in a manner which follows the liquidus curve (shown as path D in Fig. 6), whereas the T_g of the UFP moves up the glass transition line (path B; due to increased concentration) with a rapid increase in viscosity in a non-Arrhenius manner, particularly in late stages of the freezing process [12]. When a critical, solute-

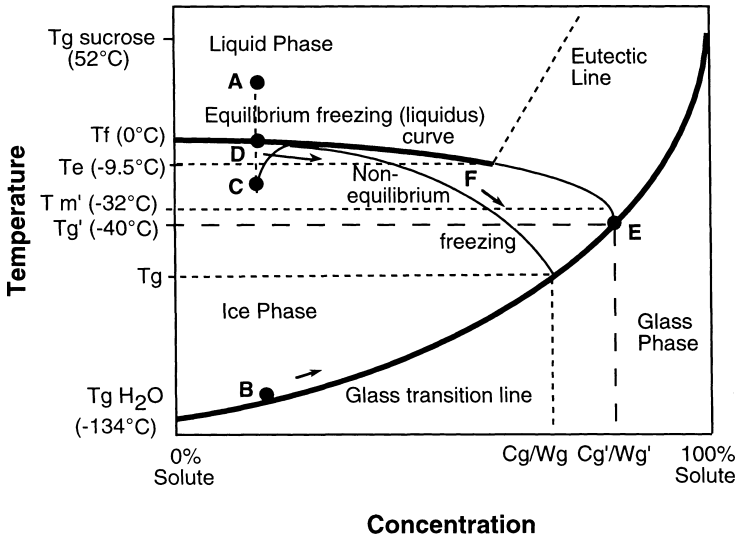


Figure 6 A schematic temperature–concentration state diagram for an aqueous carbohydrate solution, showing the glass transition curve, defined by viscosity, which extends from the T_g (glass transition temperature) of pure water (−134°C) to the T_g of pure solute, the equilibrium freezing (liquidus) curve, which extends from the T_m (melting temperature) of pure water (0°C) to the eutectic temperature (T_e) of the solute, the nonequilibrium extension of the liquidus curve from T_e to T_g', and the theoretical eutectic line. Point T_g' represents the glass transition temperature of the maximally freeze-concentrated solution at the intersection of the nonequilibrium liquidus curve and the glass transition line, and W_g' represents the amount of unfrozen water (100 minus % solute, C_g') that becomes trapped in the glass. Point T_m'reflects the temperature needed to cause mechanical collapse and ice melting to occur above the glass transition. Points T_g and W_g represent an example of a temperature concentration relationship in a glass formed as a result of less than maximal ice formation (following the "nonequilibrium" freezing line). Details of specific lettered points are provided in the text.

dependent concentration is reached, the unfrozen liquid exhibits very resisted mobility [41] and the physical state of the UFP changes from a viscoelastic liquid to a brittle, amorphous solid glass.

The intersection of the nonequilibrium extension of the liquidus curve and the glass transition curve, point E in Fig. 6, represents the solute-specific, maximally freeze-concentrated T_g of the frozen system, denoted T_g', where ice formation ceases within the timescale of normal measurement [12,42]. The corresponding maximum concentrations of water and sucrose "trapped"

within the glass at T_g' and unable to crystallize are denoted the W_g' and C_g', respectively [42–44]. It is worth noting that this unfrozen water is not bound in an "energetic" sense, rather unable to freeze within practical time frames [9,23,24]. At the T_g', the supersaturated solute takes on solid properties because of reduced molecular motion, which is responsible for the tremendous reduction in translational, not rotational, mobility [9,23,24,45,46]. It is this intrinsic slowness of molecular reorganization below T_g' that the food technologist seeks to create within the concentrated phase surrounding constituents of food materials [18,23,32,47]. However, warming from the glassy state to temperatures above the T_g' results in a tremendous increase in diffusion, not only from the effects of the amorphous to viscous liquid transition but also from increased dilution, as melting of small ice crystals occurs almost simultaneously. The onset of melting temperature of the maximally freeze-concentrated glass, T_m', may be as much as several degrees above the T_g' due to differences in viscosity attributable to molecular motion compared to macromolecular (mechanical) flow [46,48–51]. The timescale of molecular rearrangement continually changes as the T_g is approached during freezing, so that food technologists can also gain some enhanced stability at temperatures above T_g' by minimizing the ΔT between the storage temperature and T_g', either by reduced storage temperatures or enhanced T_g' through freezing methods or formulation [24,32,52,53]. Hence, knowledge of the glass transition provides a clear indication of molecular diffusion and reactivity and, therefore, shelf stability.

Despite the thermodynamic driving force to achieve the unfrozen water content corresponding to W_g', one must also consider the large kinetic factors which "overtake" the freezing process [54–56]. At subzero temperatures, the formation of an amorphous state is time dependent because the limiting factor of the process (water removal in the form of ice) becomes more difficult as concentration increases. The exponential effect of viscosity on mass transfer properties acts as the limiting factor for growth [50,51]. In addition, under conditions where heat removal is rapid, a high level of undercooling at the interface will only add to a further decrease in propagation rate [7]. The net result is that freezing becomes progressively slower as ice crystallization is hindered and, consequently, more time is required for lattice growth at each temperature [21]. Therefore, the kinetic restriction imposed on the system can lead to a situation in which nonequilibrium freezing, resulting in a partial dilute glass, can occur [54,55]. The typical pathway a system may follow during nonequilibrium freezing is shown in Fig. 6 as the line leading to a lower T_g (path F) than T_g', with a corresponding lower sucrose concentration in the glass (C_g) and higher water content in the glass (W_g) due to excess undercooled water plasticized within the glass [12]. This is often referred to as a dilute glass. The magnitude of deviation from the equilibrium curve and,

hence, the actual path followed may be regarded as a function of the degree of departure from equilibrium. Systems possessing this undesirable structure may undergo various relaxation-recrystallization mechanisms in order to maximally freeze concentrate and minimize the unfrozen water content. As a result, during warming, systems formed under these conditions may lead to one or more low-temperature transitions, followed by an exothermic devitrification peak due to crystallization of immobilized water and, finally, the onset of ice melting, T_m [54,55].

The parameters of T_g and T_g' have very distinctly different meanings, but have often been confused and there exists a wide variation in published T_g' values [32], mainly as a result of different interpretations of the same differential scanning calorimetry (DSC)-measured transitions of warming curves of many low-molecular-weight (MW) carbohydrates [53]. T_g' represents the T_g of a maximally freeze-concentrated system, whereas T_m' represents the onset of ice melting from a maximally freeze-concentrated system, which may coincide with or be a few degrees above T_g' [48,49]. Several researchers have demonstrated that proper annealing protocols within a narrow temperature range between the T_g and onset of melting, T_m, are necessary to allow for delayed crystallization and maximal freeze-concentration [54–59]. It is very difficult to form ice at concentrations >70% solute without extensive annealing (long times) and/or temperature cycling and, therefore, it is virtually impossible to maximally freeze concentrate within realistic time frames [57]. Thus, in determining T_g' values, great care must be taken to ensure maximal freeze concentration, using standardized annealing protocols. Hence, use of the term T_g' should be limited to the theoretical temperature or the measured temperature after rigorous thermal treatment to ensure maximal ice formation and freeze concentration of the solute [39,42].

Further confusion results from the typical biphasic behavior of a DSC melting thermogram and the fact that extensive annealing will cause the two transitions to approach but not to coincide. The warmer temperature transition seen in the DSC scan from $T < T_g'$ thus remains somewhat of an enigma. It has been referred to frequently as the T_m', the onset of ice melting after the glass transition [49,54,55], and its lag after the T_g is perhaps simply a reflection of the timescale of the measurement [48] or it may represent a mechanical relaxation that must occur prior to sample flow [49]. If the latter proposal is correct, then T_m' may be a more technologically significant temperature, relating to mechanical collapse and flow in the system [12].

Modulated-temperature DSC and microscopy have recently been used to examine the glass transition behavior of frozen 40% sucrose solutions under various freezing conditions [59–63] and to correlate this behavior to observed microstructural differences in the samples [60,61]. It was shown that freezing procedures that resulted in deep undercooling followed by rapid nucleation

led to two glasslike transitions upon warming [60,61]. Micrographs of the structure resulting from this freezing process revealed the presence of sucrose inclusions within the ice crystals [61], as has also been shown in freeze-concentration processes [64]. It was suggested that the glasslike nature of the second, warmer transition at T'_m was related to the presence of the solute inclusions. However, maturation of these crystals leading to a pure crystalline phase and a homogeneous bulk phase led to a change of this transition during warming from a glasslike transition to a first-order melting transition. Thus, this work showed the complexity of the glass relaxation of freeze-concentrated systems.

C. Structural Heterogeneity in Food Systems

1. Plant Tissues

Fruit and vegetable structure can be classified as cellular or tissue structure but is differentiated from animal tissues in that plants also contain cellulosic cell walls bound by a pectinaceous middle lamella. Turgor pressure is created when cell membranes exert pressure against these cell walls, giving plants their rigidity. Mature plant cells also contain a number of organelles and vacuoles. These structures are often damaged by freezing, in contrast to animal cells [17]. Loss of membrane semipermeability and disruption of cellular compartments can be minimized by a rapid freezing rate, low storage temperature, and slow thawing. Cell walls are frequently effective barriers to ice growth. Hence, on freezing, ice external to the cell will form and propagate. As heat is continually removed during freezing, the cell will be surrounded by an ice-containing matrix, which is itself becoming freeze concentrated and setting up an osmotic gradient that continually drives water out of the cell. Severely dehydrated cells can result. Upon thawing, the extracellular water drips from the tissue, causing mass loss and leaving a very undesirable, nonrigid tissue and flaccid product. With more rapid freezing, heat is removed faster than water can migrate out of the cell, causing undercooling of the cell and ice nucleation within the cell. This results in less dehydrated cells and, provided the ice crystal size is sufficiently small to prevent membrane damage, a much better product upon thawing [65–68]. Recrystallization, however, can be extremely detrimental to plant tissues, as the slow rate of refreezing during temperature fluctuation allows adequate time for moisture immigration out of the cell in the direction of the osmotic gradient [4,17]. Glass transitions occur in plant tissues at temperatures that are dictated primarily by the concentration of solutes [69].

2. Meat and Myosystems

In contrast to plant tissues, animal tissues do not contain a structurally strong cell wall. Muscle cells are flexible, elongated fibers that are aligned with

minimal air spaces. Muscle cells are less likely to break as a consequence of freezing and thawing; structural change is more likely to be evident in cell–cell separation [17]. The animal cell membrane is less effective in preventing propagation of ice and so internal freezing of animal cells is much more common than internal freezing of plant cells. Thus, in general, meat freezes well, preserves its quality, and has a very good frozen shelf life [70,71]. In most cases, the rate of freezing has very little influence on the properties of the frozen tissue and almost no effect on the quality of the thawed tissue [17]. Meat tissues can easily undergo surface sublimation [40] and lipid oxidation [72], however, if not properly packaged in oxygen- and moisture-barrier packaging.

Textural problems, when they arise, are mostly attributed to thaw rigor, protein denaturation and cross-linking, or loss in water-holding capacity. If meat is frozen after slaughter without sufficient aging, excessive drip loss and lower meat quality can result. This is attributed to thaw rigor that results from sudden high rate of ATP breakdown caused by release of Ca^{2+} from the sarcoplasmic reticulum and mitochondria. Excessive sarcomere shortening or myofibril contraction results. Freezing after completion of rigor mortis eliminates this problem. In some circumstances, changes in solute concentration can result in protein denaturation and cross-linking, leading to a soft protein exudate on thawing [5].

3. Solutions, Macromolecular Dispersions, and Gels

In nontissue systems, structural heterogeneity is less than in tissue-based systems; nevertheless, macromolecular structures of biopolymers, such as protein or polysaccharide gels, may be negatively affected by freezing. Ice formation results in the concentration of components in continually reduced spaces, therefore, a great propensity for cross-linking results. This can affect the solubility or spatial orientation. Starch retrogradation is accelerated at temperatures close to freezing [5,17]. Chemical modification to produce freeze-thaw stable starches minimizes aggregation and allows starch to rehydrate well on thawing [17]. Hydrocolloids are largely unaffected by freezing and thawing. However, locust bean gum is an example of a gum that can "cryogel"—form a stable structure due to the freeze-concentration process [73]. Hydrocolloids are often used to inhibit recrystallization in frozen foods, both those consumed frozen (e.g., ice cream) and those consumed after thawing (e.g., sauces, gravies) (see Section III.C) [28,33]. Proteins may undergo some conformational changes on freezing, often due to changes in other solution properties [e.g., salt concentration, hydrogen ion concentration (pH), ionic strength]. These affects are often reversible upon thawing [5,17]. One unique problem in eggs is yolk gelation, in which the yolk becomes hard and elastic upon thawing. This has been attributed to lipoprotein structure formation and can be ameliorated by sugar or salt addition to the yolk prior to freezing, but that limits the resulting applications of the yolk

product [5]. Emulsions generally freeze well provided the emulsion is well stabilized and the dispersed-phase volume is relatively low; otherwise, freezing can promote coalescence [5].

III. STRUCTURAL CHANGES IN FROZEN SYSTEMS

A. Shelf Life

The freezing process, the storage and distribution of frozen foods, and thawing are often each associated with a loss of quality, in comparison to an unfrozen standard. Sources of damage may be directly associated with ice formation (mechanical damage of structure), they may arise from moisture redistribution within the food product, they may result from freeze concentration of solutes directly affecting some components such as proteins or lipids, or they may result from enzymatic action. The occurrence and magnitude of any of these changes will directly affect shelf life. The rates of reactions of all of these processes are temperature dependent and perhaps more dependent on $T - T_g$ rather than absolute T [32]. In other words, whereas the critical storage temperature of vegetables might be $-18°C$, it might be more like $-30°C$ for ice cream [28]. Temperature fluctuations also greatly exacerbate most of these deteriorative reactions [4]. Thus, it has been extremely difficult for the frozen food industry to predict shelf life. It could vary from weeks to months, depending on product, packaging, and storage/distribution conditions.

As temperature is reduced, so too are rates of chemical and enzymatic reactions. Nevertheless, frozen foods are not inert; temperature decreases are offset by an increase in the concentration of reactants, due to water removal as ice, and by decompartmentalization due to ice damage, which may bring reactants together. In plants, four enzyme groups can all be active: lipoxygenase, lipase, and protease can cause off-flavor development; pectic enzymes and cellulases can cause textural changes; polyphenol oxidase, chlorophyllase, and peroxidase may cause color changes; and ascorbic acid oxidase and thiaminease can cause nutritional changes [17,65–67]. Plant materials are normally blanched prior to freezing primarily to render enzymes inactive. Blanching, however, may also cause detrimental effects in many foods [14]. In animal products, lipid oxidation may have the greatest effect on chemical deterioration. Brown discoloration due to the formation of metmyoglobin in red meats tends to parallel lipid oxidation. Airtight vacuum or nitrogen packaging to exclude oxygen and addition of antioxidants provide a measure of control [14,72].

Freezing and thawing also affect many food components. As a consequence of freeze concentration, the properties of the unfrozen phase, includ-

ing pH, titratable acidity, ionic strength, viscosity, freezing point, surface tension, and oxidation–reduction potential, are altered. These can become detrimental to food components (e.g., protein denaturation leading to curdling and drip in thawed materials). Slow freezing may destroy vegetative bacterial cells due to ice crystal damage rather than low temperature; however, due to the formation of smaller ice crystals and possibly less ice phase, rapid freezing can lead to the survival of many micro-organisms, and thus the thawing temperature becomes critical [14].

B. Moisture Migration

Moisture migration in frozen foods is responsible for moisture loss by sublimation ("freezer burn"), moisture absorption or redistribution in foods or food components, recrystallization of ice (dealt with separately), and drip loss during thawing.

Freezer burn is recognized as a change in visual appearance of frozen products at the surface. Total moisture loss by sublimation, the conversion of ice directly to the vapor phase, is only indirectly correlated to freezer burn because it is a localized surface phenomenon and its consequences depend largely on the type of food constituents found at the surface [4]. Freezer burn is also often accompanied by other deteriorative reactions such as lipid oxidation [72]. Two factors greatly affecting mass transfer associated with sublimation from the surface are barrier layers (packaging or edible barriers) and the moisture gradient between the product and air. Whereas tight packaging at the surface can protect products from sublimation, loose packaging can often lead to the undesirable appearance of frost or ice crystals on the inside of the package, as sublimated water from the product condenses on the cold inner surface of the package [4]. Moisture loss also increases with increasing storage temperature [17]. The relative humidity of storage dictates the driving force for moisture loss by sublimation. Low relative humidity increases the moisture gradient. Relative humidity is reduced by sources of dry heat within a cold store, such as electric motors, lights, and heat conduction through floors, walls, and ceilings. Moisture builds up on the coldest surfaces (e.g., the evaporation coil of the refrigeration unit). Another factor affecting moisture loss by sublimation is the surface area-to-volume ratio of the product [4]. Smaller pieces should be frozen faster, as they are more susceptible to moisture loss.

Water diffusion and redistribution during frozen storage can be another issue causing problems in frozen foods. The formation of ice changes the equilibrium moisture balance between components, causing water, usually through vapor diffusion (sublimation–condensation), to be redistributed. This occurs, for example, in multicomponent foods that have two distinct

components with differing moisture contents initially equilibrated at a constant water activity (e.g., a bakery product with a filling). Changing moisture content in one layer due to the formation of ice will affect the equilibrium distribution, usually driving moisture from the filling to the bakery product, resulting in loss of textural attributes [4]. As in nonfrozen multicomponent foods, edible barrier layers (e.g., chocolate or oil) can often be used to inhibit moisture redistribution.

Another issue arises with moisture migration during thawing. Ice crystals are pure water removed from solution or cellular structures during freezing. Upon thawing, this water may not return to the same location as it was removed from during freezing. A good example is drip loss. In tissue systems, freezing can often cause cellular dehydration, especially in plant tissues, and these cells do not significantly readsorb water. If this thawed ice is not held in extracellular spaces, drip loss occurs. Freezing regimes leading to less cellular dehydration and storage regimes that minimize recrystallization generally eliminate or minimize drip loss.

C. Ice Recrystallization

Ice crystals in frozen foods are unstable and will undergo recrystallization, the extent of which depends in part on how effectively the system has been stabilized. Recrystallization is the process of changes in the number, size, and shape of ice crystals during frozen storage, although the amount of ice stays constant with constant temperature throughout this process (dictated by the equilibrium freezing curve). Recrystallization basically involves small crystals disappearing, large crystals growing, and crystals fusing together. The driving force tends to be minimizing free energy through a decrease in surface area. This leads to greater tissue disruption and poor food quality [4]. Recrystallization can be minimized by maintaining a low and constant storage temperature.

There are several types of recrystallization process (Fig. 7). Iso-mass recrystallization ("rounding off") refers to changes in surface or internal structure so that crystals with irregular shapes and large surface-to-volume ratios assume a more compact structure. In other words, sharper surfaces are less stable than flatter ones due to the higher radius of curvature of higher specific energy and will show a tendency to become smoother over time. Migratory recrystallization refers in general to the tendency of larger crystals to grow at the expense of smaller crystals. Ostwald ripening refers to migratory recrystallization that occurs at constant temperature and pressure due to differences in surface energy between crystals, most likely involving melting–diffusion–refreezing or sublimation–diffusion–condensation mechanisms. However, migratory recrystallization is greatly enhanced by temper-

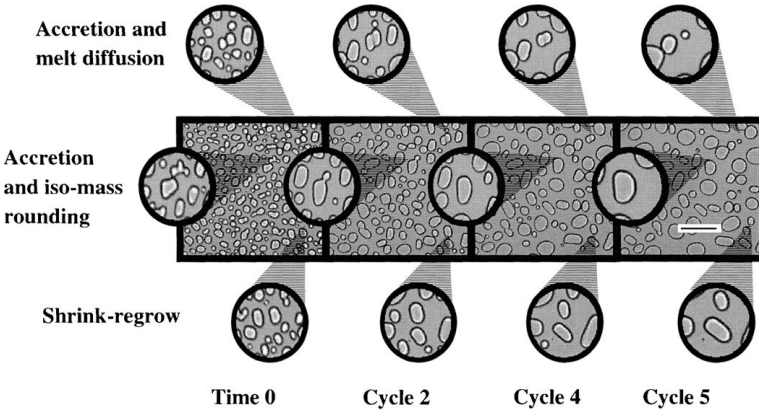


Figure 7 A sequence of light micrographs at the same subzero temperature of ice crystals in a sucrose solution after temperature fluctuation cycles showing the processes of recrystallization. Main photographs are at the same magnification (bar = 100 μm; closeup images are 5×).

ature fluctuations (heat shock) inducing a melt–refreeze behavior due to ice content fluctuations [74]. Melt–refreeze behavior can lead to complete disappearance of smaller crystals during warming and growth of larger crystals during cooling or to a decrease in size of crystals during partial melting and regrowth of existing crystals during cooling. Melt–refreeze should occur to a greater extent at higher temperatures, as more water melts and refreezes with a given ΔT due to the shape of the equilibrium freezing curve (Fig. 4) and more rapidly for smaller crystals, because they have a lower melting temperature and greater surface area [8]. As temperatures decrease during temperature fluctuation, water does not renucleate but is deposited on the surface of larger crystals, so the net result is that the total number of crystals diminish and the mean crystal size increases (Fig. 8). Accretion refers to a natural tendency of crystals in close proximity to fuse together; the concentration gradients in the areas between them are high and, therefore, material is transported to the point of contact between crystals and a neck is formed (Fig. 7). Further "rounding off" will occur because a high-curvature surface like this has a natural tendency to become planar.

Polysaccharide stabilizers (gums or hydrocolloids, e.g., guar, locust bean gum, carboxymethyl cellulose) are often used in formulated frozen foods to control ice recrystallization. A great deal of research has gone into the understanding of how they function [15,21,22,73–85]. They are often added to formulations at low levels, 0.1–0.2%. As they become freeze concentrated

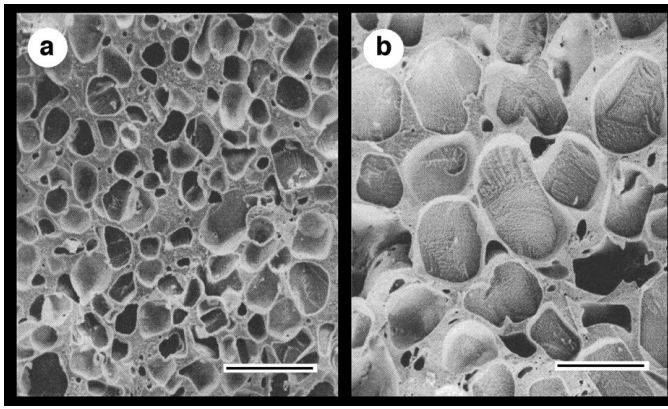


Figure 8 Cryo-scanning electron micrographs of frozen ice cream at the same magnification before (a) and after (b) temperature fluctuation cycles showing the detrimental effects of ice recrystallization on structure. Bar = 50 μm.

during freezing, however, they can exceed their critical concentration in solution and become entangled, which leads to a high viscosity in the unfrozen phase. It has also been shown that locust bean gum has the ability to gel during freezing, as a result of freeze concentration—a phenomenon known as cryo-gelation [81,83–85]. If proteins are present in the food formulation, many polysaccharides and proteins are incompatible in solution [81], leading to further localized increases in concentration and viscosity (microviscosity). The heterogeneous, gellike or porouslike structure that forms is then capable of limiting migration of water and ice during temperature fluctuations and thereby maintaining or limiting the growth of ice crystal population distributions.

There has also been recent interest in the use of ice structuring (natural "antifreeze") proteins in frozen food application [86]. These molecules are derived from plant and animal species that have the ability to overwinter. Their apparent mechanism is through direct adsorption to the surface of an ice crystal, blocking further growth along that crystal surface. As a result, ice crystal populations can be maintained during subzero temperature fluctuations.

IV. CHALLENGES FOR THE FUTURE

The frozen food industry is growing substantially and continually offering a greater array of products to meet consumer demands and increase market

share. The last decade has seen a tremendous increase in knowledge related to the stability and structure in frozen systems, particularly with respect to the recognition of the importance of the glass transition. Structure remains a very important aspect for further study, as we continually learn more about the location of ice and dislocation of existing structures and components during freezing. There still remains, however, a great number of research challenges.

We must improve our understanding of molecular mobility on both sides of the glass transition and relate this to storage stability and quality reduction reactions. We must improve our understanding of the effect of complexity of structure in real food systems on molecular glass transitions, recognizing that foods are very heterogeneous systems. We must also improve our capabilities of measuring these glass transitions in real food systems, more accurately and reliably than at present and at less complexity and cost of instrumentation. This should lead to greater predictive capabilities of shelf life and quality.

To make gains in shelf life, we need to improve our vigilant approach to maintenance of the cold chain through manufacturing, storage, distribution, retailing, and consumer handling. The frozen food manufacturer must not only recognize the importance of low constant temperature for quality of preservation, but they must educate all others involved in frozen food handling to recognize the importance of these parameters as well. However, many changes in formulation of frozen foods are also being made to improve quality preservation and make frozen foods more resistant to modulations in cold chain temperature. Polysaccharide stabilizers such as food gums or modified starches find widespread use for this application. However, undoubtedly improvements in these products and in our understanding of mechanisms of their functionality will be made. Lately, there is much interest in the ice structuring proteins extracted from biological sources. Although much research needs to be done to understand their functionality and adapt them to food applications, the upcoming decade will undoubtedly also see a large increase in their use.

REFERENCES

1. Mallett, C.P., Ed.; *Frozen Food Technology*; Blackie Academic and Professional: New York, 1993.
2. Reid, D.S. In *Quality in Frozen Food*; Erickson, M.C., Hung, Y.C., Eds.; Chapman & Hall: New York, 1997; 10–28.
3. *Quality in Frozen Food*; Erickson, M.C., Hung, Y.C., Eds.; Chapman & Hall: New York, 1997.
4. Pham, Q.T.; Mawson, R.F. In *Quality in Frozen Food*; Erickson, M.C., Hung, Y.C., Eds.; Chapman & Hall: New York, 1997; 67–91.

5. Haard, N.F. In *Quality in Frozen Food*; Erickson, M.C., Hung, Y.C., Eds.; Chapman & Hall: New York, 1997; 275–295.
6. Jeremiah, L.E., Ed.; *Freezing Effects on Food Quality*; Marcel Dekker: New York, 1976.
7. Franks, F. In *Water: A Comprehensive Treatise*; Franks, F., Ed.; Plenum Press: New York, 1982; Vol. 7, 215–338.
8. Franks, F. *Biophysics and Biochemistry at Low Temperatures*; Cambridge University Press: Cambridge, 1985; 21–61.
9. Franks, F. *Biophysics and Biochemistry at Low Temperatures*; Cambridge University Press: Cambridge, 1985; 167–187.
10. Garside, J. In *Food Structure and Behaviour*; Blanshard, J.M.V., Lillford, P., Eds.; Academic Press: New York, 1987; 35–50.
11. Roos, Y. *Phase Transitions in Foods*; Academic Press: New York, 1995; 1–360.
12. Sahagian, M.E.; Goff, H.D. In *Freezing Effects on Food Quality*; Jeremiah, L.E., Ed.; Marcel Dekker: New York, 1976; 1–50.
13. Reid, D.S. Food Technol. 1983, *37* (*4*), 110–115.
14. Reid, D.S. Food Technol. 1990, *44* (7), 78–82.
15. Muhr, A.H.; Blanshard, J.M.V.; Sheard, S.J. J. Food Technol. 1986, *21*, 587–603.
16. Flores, A.A.; Goff, H.D. J. Dairy Sci. 1999, *82*, 1399–1407.
17. Reid, D.S. In *Frozen Food Technology*; Mallett, C.P., Ed.; Blackie Academic and Professional: New York, 1993; 1–19.
18. Blanshard, J.M.V.; Franks, F. In *Food Structure and Behaviour*; Blanshard, J.M.V., Lillford, P., Eds.; Academic Press: New York, 1987; 35–50.
19. Leighton, A. J. Dairy Sci. 1927, *10*, 300–308.
20. de Cindio, B.; Correra, S.; Hoff, V. J. Food Eng. 1995, *24*, 405–415.
21. Muhr, A.H.; Blanshard, J.M.V. J. Food Technol. 1986, *21*, 683–710.
22. Blond, G. Cryobiology 1988, *25*, 61–66.
23. Levine, H.; Slade, L. Cryo-Letters 1988, *9*, 21–60.
24. Slade, L.; Levine, H. CRC Crit. Rev. Food Sci. Nutr. 1991, *30*, 115–360.
25. Blond, G.; Simatos, D. Thermochimi Acta. 1991, *175*, 239–247.
26. Levine, H., Slade, L., Eds.; *Water Relationships in Foods*; Plenum Press: New York, 1991.
27. Blond, G.; Simatos, D. Thermochim. Acta 1991, *175*, 239–247.
28. Goff, H.D. Food Res. Int. 1992, *25*, 317–325.
29. Blanshard, J.M.V., Lillford, P.J., Eds.; *The Glassy State in Foods*; Nottingham University Press: Loughborough, 1993.
30. Simatos, D.; Blond, G. In *Glassy State in Foods*; Blanshard, J.M.V., Lillford, P., Eds.; Nottingham University Press: Loughborough, 1993; 395–416.
31. Goff, H.D. Pure Appl. Chem. 1995, *67*, 1801–1808.
32. Slade, L.; Levine, H. In *Advances in Food and Nutrition Research*; Kinsella, J.E., Taylor, S.L., Eds.; Academic Press: New York, 1995; Vol. 38, 103–269.
33. Goff, H.D.; Sahagian, M.E. Thermochim. Acta 1996, *280*, 449–464.
34. Goff, H.D. In *Quality in Frozen Food*; Erickson, M.C., Hung, Y.C., Eds.; Chapman & Hall: New York, 1997; 29–50.

35. Levine, H., Ed.; *Amorphous Food and Pharmaceutical Systems*; Royal Society of Chemistry: Cambridge, 2002.
36. Angell, C.A.; Choi, Y. J. Microscopy 1986, *141*, 251–261.
37. Angell, C.A.; Bressel, R.D.; Green, J.L.; Kanno, H.; Oguni, M.; Sare, E.J. J. Food Eng. 1994, *22*, 115–142.
38. Blond, G. Cryo-Letters 1989, *10*, 299–308.
39. Ablett, S.; Clark, A.H.; Izzard, M.J.; Lillford, P.J. J. Chem. Soc. Faraday Trans. 1992, *88*, 795–802.
40. Blond, G.; Simatos, D.; Catte, M.; Dussap, C.G.; Gros, J.B. Carbohydrate Res. 1997, *298*, 139–145.
41. Carrington, A.K.; Sahagian, M.E.; Goff, H.D.; Stanley, D.W. Cryo-Letters 1994, *15*, 235–244.
42. Matveev, Y.I.; Ablett, S. Food Hydrocoll. 2002, *16*, 419–422.
43. Hatley, R.H.M.; van den Berg, C.; Franks, F. Cryo-Letters 1991, *12*, 113–124.
44. Hatley, R.H.M.; Mant, A. Int. J. Biol. Macromol. 1993, *15*, 227–232.
45. Le Meste, M.; Huang, V. J. Food Sci. 1992, *57*, 1230–1233.
46. Blond, G. J. Food Eng. 1994, *22*, 253–269.
47. Simatos, D.; Blond, G. In *Water Relationships in Foods*; Levine, H., Slade, L., Eds.; Plenum Press: New York, 1991; 139–156.
48. MacInnes, W.M. In *Glassy State in Food*; Blanshard, J. Lillford, P., Eds.; Nottingham University Press: Loughborough, 1993; 223–248.
49. Shalaev, E.Y.; Franks, F. J. Chem. Soc. Faraday Trans. 1995, *91*, 1511–1517.
50. Sun, W.Q. Cryo-Letters 1997, *18*, 99–106.
51. Cruz, I.B.; Oliveira, J.C.; MacInnes, W.M. Int. J. Food Sci. Technol. 2001, *36*, 539–550.
52. Lim, M.H.; Reid, D.S. In *Water Relationships in Foods*; Levine, H., Slade, L., Eds.; Plenum Press: New York, 1991; 103–122.
53. Reid, D.S.; Kerr, W.; Hsu, J. J. Food Eng. 1994, *22*, 483–494.
54. Roos, Y.; Karel, M. Int. J. Food Sci. Technol. 1991, *26*, 553–566.
55. Roos, Y.; Karel, M. Cryo-Letters 1991, *12*, 367–376.
56. Sahagian, M.E.; Goff, H.D. Thermochim. Acta 1994, *246*, 271–283.
57. Ablett, S.; Izzard, M.J.; Lillford, P.J. J. Chem. Soc. Faraday Trans. 1992, *88*, 789–794.
58. Ablett, S.; Izzard, M.J.; Lillford, P.J.; Arvanitoyannis, I.; Blanshard, J.M.V. Carbohydrate Res. 1993, *246*, 13–22.
59. Aubuchon, S.R.; Thomas, L.C.; Theuerl, W.; Renner, H. J. Thermal. Anal. 1998, *52*, 53–64.
60. Goff, H.D.; Montoya, K.; Sahagian, M.E. In *Progress in Amorphous Food and Pharmaceutical Systems*; Levine, H., Ed.; Royal Society of Chemistry: Cambridge, 2002; 145–157.
61. Goff, H.D.; Verespej, E.; Jermann, D. Thermochim. Acta 2003, *399*, 43–55.
62. Izzard, M.J.; Ablett, S.; Lillford, P.J.; Hill, V.L.; Groves, I.F. J. Thermal. Anal. 1996, *47*, 1407–1418.
63. Knopp, S.A.; Chongprasert, C.; Nail, S.L. J. Thermal. Anal. 1998, *54*, 659–672.
64. Chen, P.; Chen, X.D.; Free, K.W. J. Food Eng. 1998, *38*, 1–13.

65. Skrede, G. In *Freezing Effects on Food Quality*; Jeremiah, L.E., Ed.; Marcel Dekker: New York, 1976; 183–246.
66. Cano, M.C. In *Freezing Effects on Food Quality*; Jeremiah, L.E., Ed.; Marcel Dekker: New York, 1976; 247–298.
67. Arthey, D. In *Frozen Food Technology*; Mallett, C.P., Ed.; Blackie Academic and Professional: New York, 1993; 237–269.
68. Allan-Wojtas, P.; Goff, H.D.; Stark, R.; Carbyn, S. Scanning 1999, *21*, 334–347.
69. Sá, M.M.; Sereno, A.M. Thermochim. Acta 1994, *246*, 285–297.
70. Marsden, J.L.; Henrickson, R.L. In *Frozen Food Technology*; Mallett, C.P., Ed.; Blackie Academic and Professional: New York, 1993; 168–195.
71. Devine, C.E.; Bell, R.G.; Lovatt, S.; Chrystall, B.B.; Jeremiah, L.E. In *Freezing Effects on Food Quality*; Jeremiah, L.E., Ed.; Marcel Dekker: New York, 1976; 51–84.
72. Erickson, M.C. In *Quality in Frozen Food*; Erickson, M.C. Hung, Y.C., Eds.; Chapman & Hall: New York, 1997; 141–173.
73. Harper, E.K.; Shoemaker, C.F. J. Food Sci. 1983, *48*, 1801–1803.
74. Budiaman, E.R.; Fennema, O. J. Dairy Sci. 1987, *70*, 534–546.
75. Budiaman, E.R.; Fennema, O. J. Dairy Sci. 1987, *70*, 547–554.
76. Buyong, N.; Fennema, O. J. Dairy Sci. 1988, *71*, 2630–2639.
77. Sahagian, M.E.; Goff, H.D. Food Hydrocoll. 1995, *9*, 181–188.
78. Sahagian, M.E.; Goff, H.D. Food Res. Int. 1995, *28*, 1–8.
79. Carrington, A.K.; Goff, H.D.; Stanley, D.W. Food Res. Int. 1996, *29*, 207–213.
80. Flores, A.A.; Goff, H.D. J. Dairy Sci. 1999, *82*, 1408–1415.
81. Goff, H.D.; Ferdinando, D.; Schorsch, C. Food Hydrocoll. 1999, *13*, 353–364.
82. Martin, D.R.; Ablett, S.; Darke, A.; Sutton, R.L.; Sahagian, M. J. Food Sci. 1999, *64*, 46–49.
83. Regand, A.; Goff, H.D. J. Dairy Sci. 2002, *85*, 2722–2732.
84. Regand, A.; Goff, H.D. Food Hydrocoll. 2003, *17*, 95–102.
85. Patmore, J.V.; Goff, H.D.; Fernandes, S. Food Hydrocoll. 2003, *17*, 161–169.
86. Goff, H.D.; Regand, A.; Tharp, B.W. Dairy Ind. Int. 2003, *67* (10), 30–32.

13
Biogenic Cellular Solids

MARTIN G. SCANLON University of Manitoba, Winnipeg, Manitoba, Canada

I. INTRODUCTION

A. Classification

Cellular solids are one category of soft solids that are found in various forms throughout nature and the food chain. Much of the credit for rationalization of the diverse groups of materials within this category is due to Gibson and Ashby [1,2], whose approach of combining "simple mechanics with scaling ideas" has greatly simplified our understanding of their structure–property relationships. Their book, *Cellular Solids*, has been described as "a classic" [3] and has been well cited in the biological sciences as well as in the engineering literature.

Although cellular solids are frequently indexed separately from foams, they are actually a subset of them, comprised of a solid network associated with a fluid phase. Due to the continuity of the solid phase, cellular solids are stable over reasonable periods of time, in contrast to the inherent instability of liquid foams [4]. However, because the primary means of fabricating cellular solids is by the setting (rigidification) of a liquid foam, issues of stability in liquid foam structures [3] have a bearing on the structure, and hence the properties, of cellular solids [5,6].

Cellular solids are also a subset of porous solids (if the fluid phase is air) or saturated porous solids (if the fluid phase is liquid). The distinguishing feature for cellular solids is their high porosity (volume fraction >70%), a porosity cutoff that ensures that cell wall thickness relative to cell wall length is small, permitting simplification of the analysis of the mechanical properties [2]. The utility of cellular solids' residence within this subset of materials is that rigorous bounds on many physical properties of the cellular solids can be defined [7]. This is beneficial from a food processing perspective in that it may be desirable to know the extent to

which a parameter such as the modulus, which often correlates well with sensory assessements of texture [8], can be manipulated by changes in the properties of the solid phase or the porosity. The latter parameters are, in turn, manipulable by ingredient formulation or by altering process variables.

B. Regular Cellular Structures

The basic repeating unit in a cellular solid is the cell. Packing of the cells gives rise to the structure of the cellular solid. A very efficient packing structure [9] is comprised of a periodic assembly of 14-sided (tetrakaidecahedral) cells (Fig. 1). The lower the volume fraction of the solid, the greater the length–thickness ratio of the walls. The structures created from the cells of Fig. 1a or 1b represent a further gross distinction in the classification of cellular solids: intersecting lines or planes of solid around cells of fluid giving rise to open- or closed- cell cellular solids, respectively. This distinction can cause substantial differences in the properties for cellular solids of identical solids volume fraction and identical solid-phase material. As an example, the diffusion of components out of a cellular solid will be very much faster in an open-cell cellular solid compared to its closed-cell counterpart.

In order to understand how different structural arrangements affect the physical properties of cellular solids, a useful ploy is to reduce the dimensionality of the cellular solid [2,3]. By performing experimental and theoretical analyses on cellular solids of less complicated geometry, the mechanisms governing the physical properties of three-dimensional cellular solids can be more easily identified [2]. The honeycomb is a common two-dimensional

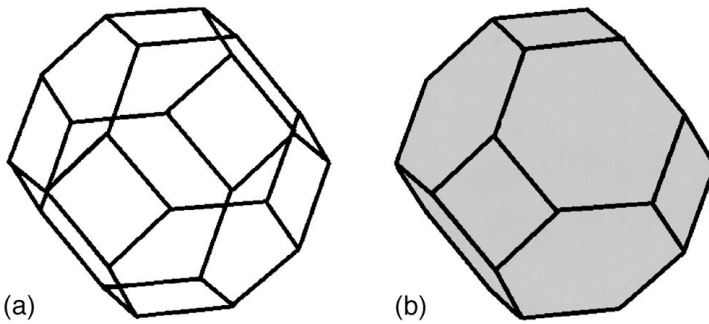


Figure 1 The 14-sided (tetrakaidecahedral) Kelvin cell, as an open-cell (a) and closed-cell (b) structure.

structure (Fig. 2), for which one physical property, its density (ρ^*), can be readily calculated from geometrical considerations:

$$\rho^* = \rho_f + \rho_s\left(1 - \frac{\rho_f}{\rho_s}\right)\left(\frac{2t}{\sqrt{3}l} - \frac{t^2}{3l^2}\right) \tag{1}$$

Equation (1) reduces to Eq. (2) when the solid density, ρ_s, is much greater than ρ_f (the fluid density), as is the case in air-filled honeycombs,

$$\rho^* = \rho_s\left(\frac{2t}{\sqrt{3}l} - \frac{t^2}{3l^2}\right) \tag{2}$$

and to Eq. (3) when $l \gg t$,

$$\rho^* \approx \rho_s\left(\frac{2t}{\sqrt{3}l}\right) \tag{3}$$

In the complicated geometry of a three-dimensional cellular solid, deriving an exact expression for density from the structure would be challenging.

C. The Softness of Cellular Solids

By their very definition of substitution of a portion of the solid with a fluid, one expects cellular solids to exhibit reduced mechanical properties compared to those of the solid that comprises the solid network, regardless of the solid's structural organization. This "softening" is apparent from the expression used to define a property, such as the elastic modulus, E^*, of a cellular solid [10]:

$$E^* = E_s C_1 \varphi^m \tag{4}$$

where E_s is the modulus of the solid material comprising the cell wall, C_1 is a constant, and φ is the volume fraction of solid material, which is equivalent to the relative density (ρ^*/ρ_s) when the cellular solid is gas-filled.

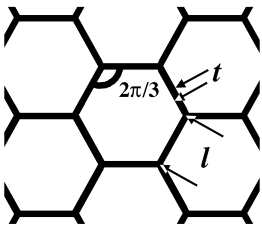


Figure 2 A honeycomb structure, where the structural characterization parameters t and l represent the thickness and cell edge length of cell walls, respectively.

The yield stress, σ_y^*, under compression is similarly defined, so that for a gas-filled cellular solid,

$$\sigma_y^* = \sigma_{ys} C_2 \varphi^n = \sigma_{ys} C_2 \left(\frac{\rho^*}{\rho_s} \right)^n \tag{5}$$

where σ_{ys} is the yield stress of the solid material and C_2 is a constant. Theoretical values for the exponents m and n, based on specific structural arrangements of the solid and on specific modes of deformation, have been summarized by Lakes [11]. Experimental values ranging between 1.0 and 4.3 have been found for m for starch-based gas-filled cellular solids made by baking or extrusion [12,13]. Therefore, for a relative density of 0.15, one expects from Eq. (4) that the modulus of a cellular solid might only be 0.03% of the modulus of the solid from which it is fabricated. Accordingly, biogenic cellular solids are doubly soft–soft from the effect of interactions that typically operate at much larger length scales in biogenic solids [14] and soft from substitution of this soft solid with a fluid. The softening can be mitigated somewhat when the solid phase is organized into a closed-cell structure or when loading rates are high for open-cell structures, particularly where the fluid is liquid [2,15]. Whether a cellular solid has an open- or closed-cell structure depends largely on how Nature creates her cellular solids or how humanity uses Nature's materials to create specific cellular structures through industrial or domestic processing operations.

II. FORMATION

A. Biochemical Self Assembly

The two main categories of biogenic cellular solids that are constructed by Nature are cancellous bone ($0.05 \leq \varphi \leq 0.7$ [2]) and diverse tissues fabricated from the structural organization of plant cells ($0.05 \leq \varphi \leq 0.8$ [2]; in parenchymatous tissue, the main source of edible plant foods, $\varphi \sim 0.1$ [16]). The cell walls of both cellular solids possess an extremely complicated structure, and much remains to be learned of their mechanisms of formation [2,17,18]. Even when considering one polymer (cellulose) that figures prominently as a component of plant cell walls, a complex architecture is apparent [19]. Properties of plant cell walls are therefore highly variable [18], dependent on the effects of various environmental factors that prevail during cell wall formation. One primary factor influencing structural organization is stress induced by external forces such as gravity and wind [18]. Stress also plays a role in the formation of cancellous bone [2], with persistent stress required to maintain the bone's relative density and thus prevent the initiation of fractures.

B. Processing

The majority of man-made cellular solids are created from liquid foams. Therefore, during processing, a new phase must be nucleated in the form of gas cells (bubbles). Nucleation is sometimes followed by bubble growth. The inherently unstable foam is then converted into a cellular solid by rigidification of the liquid foam matrix into a solid matrix, albeit a viscoelastic solid for most foods. Rigidification maintains the cellular structure throughout the shelf life of the processed food. Density data for some typical biogenic cellular solids are compiled in Table 1.

1. Nucleation

There are three basic means of creating the nuclei that will form the gas cells of the foam [2,3,33]. Gas can be introduced deliberately by a technique such as

Table 1 Some Selected Biogenic Cellular Solids and Their Densities

Cellular solid	Density (ρ^*) (kg/m^3)	Solid density (ρ_s) (kg/m^3)	Volume fraction (φ)	Ref.
Apple	790–880			20
Bread (commercial)	130–145	1542 ± 29	0.08–0.09	12, 21
Breakfast cereals				
Extruded rings	150–240	1440–1500	0.1–0.17	22
Puffed grains	80–220	1390–1540	0.06–0.15	22
Shredded maize	200–210	1510	0.14	22
Shredded wheat	390–470	1440–1460	0.23–0.32	22
Cake	310	1650	0.19	23
	180–290			24
Cancellous bone	1800–2000		0.05–0.7	2
Cork	170	1150	0.15	2
Dried confections	100–300			25
Freeze-dried bread	80–280	1430–1620	0.05–0.2	12, 26
Marshmallow			0.17–0.29	27
Mousse (high overrun)			0.22–0.33	28
Parenchyma			~0.1	16
Potato parenchyma	1030–1170			29
Rice cakes	115–185			30
Rice snacks (extruded)	170–300	1550	0.1–0.2	31
Wheat crisp bread (≤30% bran)	85–240			32
Woods		1500		2
Hardwoods	200–750		0.13–0.5	2
Softwoods	350–590		0.23–0.4	2

sparging, so that gas enters the liquid from frits of uniform size, ensuring that cell sizes are essentially monodisperse [3]. A typical food product whose gas cells are nucleated in this manner is mousse [28]; typical nucleating gases are air, nitrogen, or nitrous oxide; the latter two are desirable when oxygen has deleterious effects on the properties of the cellular solid. Second, gas can be entrained into the liquid by mechanical energy, most commonly by a gas being present at a given pressure in the headspace of a mixing vessel. Subdivision of the entrained gas bubbles generates a distribution of gas nuclei throughout the liquid. The gas phase in bread dough is typically nucleated in this manner, and although the entrained gas is usually air, different gases can create markedly different gas cell structures in the foam, which, in turn, translate into cellular solids of different structures [34]. As in comminuton, repetitive size reductions of the entrained gas bubbles lead to a wide distribution in nucleus sizes [35]. The third means of getting gas cells to nucleate is to create a supersaturated "solution" by altering temperature and/or pressure; the supersaturated gas nucleates gas cells in the classical sense of nucleation. Typical gases are CO_2 or water vapor.

2. Growth of Nuclei into Gas Cells

This stage may be omitted for some cellular solids, so that, ideally, no change in gas cell structure occurs between nucleation and rigidification. An example is angel food cake batter, where the gas cells are formed by mixing a batter, and baking occurs as soon as possible thereafter to prevent a change in gas cell structure. For other foods, particularly high viscosity liquids such as extrudate and bread doughs, the original nuclei grow into larger gas cells. Growth is due to further expansion of the cells by the gas that was supersaturated prior to nucleation or to generation of gas by biological or chemical blowing agents within the liquid matrix. In addition to the mean gas cell size increasing with time, the size distribution coarsens due to the effects of a number of physical factors [36].

Where gases are generated in the liquid matrix, the gases are generally not those used to nucleate the bubbles, the most common ones being carbon dioxide and water vapor. Carbon dioxide may be generated by biological means (e.g., yeast or bacteria) or it may be generated by chemical leavening agents. In the latter case, control of the rate of CO_2 generation, and therefore the rate of gas cell growth, can be manipulated by the acid and bicarbonate salt type [37]. This introduces other variables for manipulating gas cell growth rates in addition to temperature which is used in biological leavening. Steam is another gassing agent, generated from water in the matrix. This technique of gas cell growth is commonly used to expand certain snack products, such as those made from potato starch doughs. Low-temperature extrusion creates nuclei in the moist dough matrix of the "half-product" from air incorporated within the dry ingredients. Immersion of the half-products in hot oil (>100°C)

generates steam from the moisture within the dough and thus expands the gas cells into a low-density snack product.

A further means of gas cell growth is simply that associated with Boyle's law. The foam is heated up and the gases expand, provided that the viscosity of the liquid phase is low enough to accommodate gas cell growth. Growth of gas cells by thermal expansion can be augmented by evaporation of CO_2 and water vapor from the liquid matrix into the gas cells as their vapor pressures rise with temperature [38].

One factor that affects gas cell structure as a result of the time required for gas cell growth is drainage of the liquid [3,5,6]. Liquid moves from cell faces to the lower-energy-state cell edges, and then, for low-viscosity liquids, it travels downward through the foam. Stabilization of the faces against bursting allows closed-cell cellular solids to be created from the liquid foam and open-cell solids to form where faces are not stabilized. Where gas cell growth occurs and the viscosity of the liquid is high, such as in bread doughs, the rheology of the liquid as well as its surface tension can prevent cell faces from bursting [36] and therefore improve the quality of the resulting cellular solid.

3. Rigidification

The conversion of liquid foam into a cellular solid is accomplished by the addition or removal of heat. Where heat is removed, rigidification may arise from sufficient enhancement of viscosity that the foam solidifies. These cellular solids remain stable provided the temperature remains low [39]. In addition to the temperature drop, desiccation of the foam may enhance stability, so that the cellular solid is no longer as sensitive to thermal abuse [39]. Solidification of a component within the liquid matrix of the foam can also be a means of creating a cellular solid, e.g., aerated chocolate bars, where fat solidifies, or some high-overrun frozen desserts, where water does.

Where heat is added to the liquid foam, rigidification arises from chemical and/or physical changes in the properties of the foam or from compositional changes, especially desiccation. Some of the changes arise from the reactivity of functional groups on polymers leading to the formation of chemical cross-links [40]. Other foams are rigidified by physical cross-links [e.g., bread, where leaching of long-chain molecules (amylose) causes a sintering of partially gelatinized starch granules, a phenomenon that was not feasible when the granules were structurally sound at lower temperatures] [41].

III. STRUCTURE

A. Defining Structure

As with other soft materials, structure plays a very significant role in the properties of cellular solids. Their most readily identifiable structural feature

is the cell (Figs. 1 and 2). Relationships between the dimensions of cell and cell wall determine the relative density—"the single most important structural characteristic of a cellular solid" [1,2]. Despite the strength of this assertion, Gibson and Ashby devoted more space to consideration of the effects of additional structural parameters in the second edition of their book [2], particularly with respect to collapse or yielding phenomena. Examples of these additional structural parameters are irregularity in cellular structure [42], anisotropy in the cells [43], defects in the cellular assembly due to missing cell walls [44,45], and irregularities in cell wall shape [6] and thickness [5]. Investigations of the effects of these additional structural features are particularly pertinent because the origins of structure in the vast majority of food cellular solids are attributable to random biochemical and/or physical processes and because the materials comprising the solid part of the cell wall have complex structural hierarchies in even "simple" foods [46,47]. An illustration of the various length scales that may be present in the solid phase of a cellular solid is given in Fig. 3 for plant parenchyma tissue. Similar scaling is apparent in bread crumb [47] and in bone [11]. A hierarchy of structures within structures is evident, and each level of microstructure influences the properties of the solid component [48], which, as is evident from Eqs. (4) and (5), are key parameters defining the properties of the cellular solid. For example, in Fig. 3, the orientation and the amount of cellulose microfibrils strongly determines the modulus of the cell walls, thereby affecting the modulus of the parenchyma [18,19,49].

The fluid can also display a hierarchical distribution, so that cell walls surrounding fluid can themselves be punctuated by pockets of fluid, leading to an additional level of cellularity within the cell wall. This additional level of structure modifies the cellular solid's properties, with the extent of modification dependent on the property being measured and whether additional hierarchical levels exist [11,48]. If these levels of hierarchical ordering extend over a sufficiently large number of length scales, then fractal concepts can be invoked to characterize the cellularity [50,51]. As described in other chapters, the parameters that quantify such a distribution can be used to predict various properties of the solid as a whole; in cellular solids, predictions have been reported for electrical conductivity [51] and elastic modulus [41,51].

B. Measuring Structure

Structure in cellular solids can be quantified by cutting a number of sections in the solid and acquiring images of the surfaces of the cross sections. Image analysis can then be used to threshold the image into cell walls and fluid (open space). From a knowledge of the density of the two phases, the relative density in two dimensions can be calculated [21]. By zooming in on the cell wall using

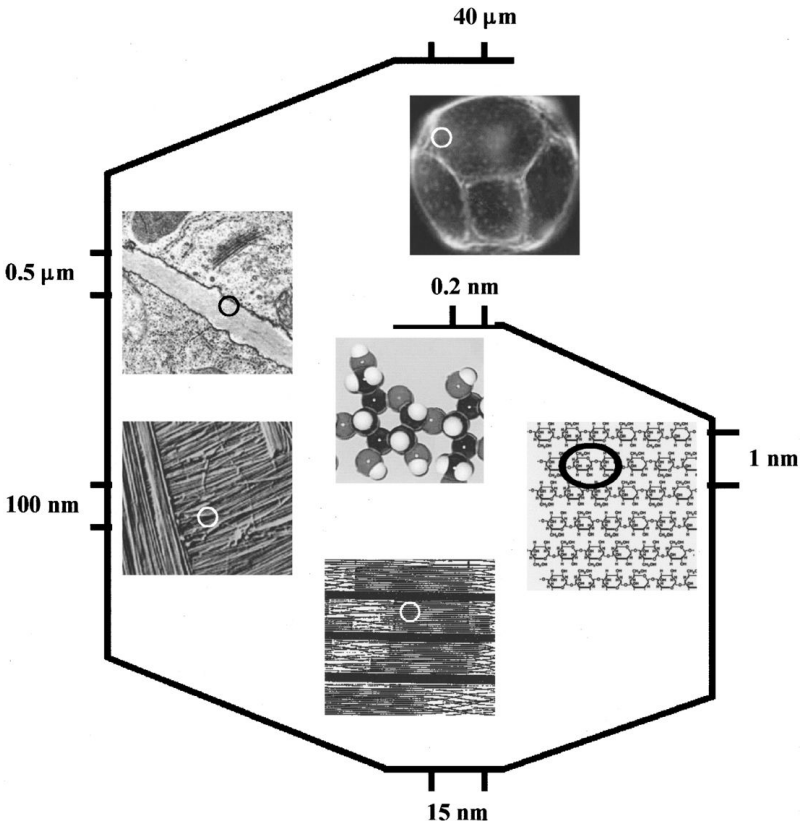


Figure 3 Schematic of the hierarchy of structures evident in going from glucose to plant cell. [Individual pictures courtesy of ©SPIRING ENTERPRISES LTD, ENGLAND (molymod.com), Professor R. Malcolm Brown, Jr. (www.botany.utexas.edu), Professor Gunning, Professor Steer and Arnold Hodder (London), Institute of Food Research (Norwich) and Professor K. Waldron.]

light [21] or electron [52,53] microscopic techniques, structure can be quantified at smaller length scales. Repeating this quantification at a number of length scales allows a fractal dimension to be extracted for both phases [41,50].

Two issues arise from taking such an approach. The first is the validity of deriving information on a three-dimensional structure from two-dimensional images. Fortunately, as shown by Underwood [54], the volume fraction of the fluid phase is equivalent to its areal fraction. The second issue is definition of a threshold that allows accurate demarcation of solid and fluid phases. Scanlon

and Zghal [55] showed that the number of cells detected in images of bread crumb fell sharply either side of a specific gray-level threshold which maximized areal cell density. Therefore, unless rigorous objective strategies are employed to govern threshold setting, the likelihood is that the values of relative density obtained from image analysis of any one given cellular solid will vary widely [56]. It follows from this that lighting (and its uniformity) across the cross section being imaged is critical to the attainment of an accurate and global threshold in the first place.

A technique that is becoming accessible for quantifying structure in cellular solids is three-dimensional imaging, or tomography [57]. Tomographic images of the cellular structure of bone are shown in Fig. 4. This technique has the advantage that cell structure can be quantified directly *in situ* without the need for invasive cross sectioning. In addition, it does not rely on the assumptions inherent in the conversion of two-dimensional densities to their three-dimensional counterparts. Internal two-dimensional cross sections can be generated from the tomographic image. The issue of objective thresholding is still pertinent to tomography to ensure accurate measurement of relative densities, and elaborate procedures have been employed so that the spatial distribution of cell walls is accurately represented [57].

An obvious application of tomography is to evaluate whether the insights gained on the modes of deformation and collapse of honeycombs [44] are actually relevant to deformation mechanisms in three-dimensional cellular solids. Elliot et al. [57] showed that, just as honeycombs do, three-dimensional cellular solids develop localized regions of densification after a critical stress is surpassed. Such evidence strongly supports the view that, rather than all of the cells collapsing globally [1], seemingly minor perturbations in structural uniformity exert a powerful influence on the collapse of cellular solids.

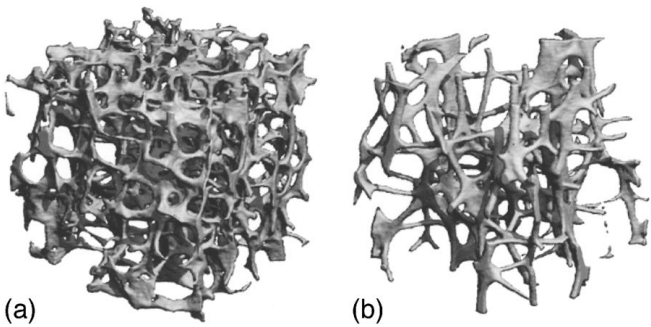


Figure 4 Tomographs of normal (a) and osteoporitic (b) cancellous bone of human vertebrae. [Courtesy of SCANCO Medical AG, Bassersdorf, Switzerland.]

IV. MODELING STRUCTURE: PROPERTY RELATIONSHIPS

A. Properties

For industrial cellular solids, a number of properties are useful, particularly mechanical and transport properties. These properties determine whether the cellular solid will perform adequately when used, for example, as a protective package or as an insulator in an energy-efficient home. For biogenic cellular solids, mechanical and transport properties are no less important (e.g., load-bearing capacity and transport of nutrients in bones), but, in addition, for cellular solids that are to be consumed as food, optical properties must also be considered.

B. Optical Properties

The nature of the interactions of light with the biogenic cellular solid affects our perception of its appearance, and because food appearance is an important quality parameter [58], optical properties influence consumer acceptability. Some cellular solids possess a glossy outer surface, typically due to a higher relative density near the surface arising from interaction of the liquid foam with equipment surfaces during processing. The gloss is attributable to enhanced specular reflection at the surface. A cellular solid without such a modified surface interacts with light in much the same way as its bulk does. The obvious qualification is that the surface is not flat, with variation in surface height of the order of half the mean cell size, and this surface roughness will lead to substantial diffuse reflection. In the bulk of the cellular solid, both scattering and absorption of the light must be considered in an examination of factors affecting optical properties.

Absorption of light will evidently depend on the molecular structure of the ingredients that constitute the solid phase of the cellular solid. Thus, the addition of egg yolks to a cake mixture will lead to selective absorption of the light by the xanthophyll as the light strikes the cell walls, so that the light that is diffusely reflected from, or transmitted through, the walls of the cellular solid will be yellow [58,59]. The extent of this absorption effect will depend on the concentration of xanthophyll in the eggs added to the mix [58]. For those cellular solids where the fluid phase is liquid, the optical absorptivity of the liquid will contribute to the optical absorption of the cellular solid as a whole. Thus, beets (beetroots), with significant amounts of betalains in the vacuoles, absorb light strongly in the blue and green regions of the spectrum, giving rise to a purple color in the cellular structure.

Scattering of light, rather than absorption, is often an effective means of altering the appearance of a cellular solid. The magnitude of scattering in

cellular solids depends on two relative ratios: the ratio of the size of the scattering feature to the wavelength of the light and the ratio of the refractive indices of the solid and fluid phases. Therefore, scattering will be most effective when cell wall thickness is of the order of 0.5 μm and when the fluid phase is gas, because for most materials in biogenic structures, the ratio of refractive indices is maximized for air rather than for aqueous solutions [60]. The ability to modify the refractive index of the fluid phase by dissolution of components further attenuates the contrast between cell wall and cell in liquid-filled cells compared to their gaseous counterparts. Typical sizes of the primary cell wall in plants range from 0.1 to 1.5 μm [19], so that substantial scattering of light is expected from the cell walls. Good illustrations of the effect of cellularity on the diffusion of light (taking into account the combined effects of scattering and absorption) are given by Dahm and Dahm [59].

Apple parenchyma represents an interesting cellular solid where differences in the ratios of refractive index can be exploited to detect a physiological disorder known as water core [61]. Apple parenchyma is an anisotropic cellular solid of liquid-filled cells, but with a significant proportion (>20%) of "cells" comprised of air (extracellular air spaces). In apples affected by water core, the gas-filled cells become filled with liquid and the extent of scattering in the cellular assembly is diminished due to the reduction in the ratio of refractive indices. Light scattering can also be exploited to evaluate the size of cells in gas-filled cellular solids [4].

In other regions of the electromagnetic spectrum, scattering and absorption of the electromagnetic radiation are also relevant, and they have a bearing on other properties of the cellular solid. For biogenic cellular solids, where water is frequently a major constituent, either within the solid or the fluid phase, strong absorption bands in the infrared increase the proportion of radiation that is absorbed relative to that scattered.

C. Mechanical Properties

Because mechanical properties determine textural quality attributes, as well as load-bearing or energy-absorbing capabilities, the static mechanical properties of cellular solids have been extensively studied. The density scaling relationships of Gibson and Ashby have been used to characterize and predict various mechanical properties of biogenic cellular solids, including those of extrudates [13,62–64], wood [2], cork [2,65], cake [24], bone [2], bread [12,41,55], starch bread [66], freeze-dried bread [26], and freeze-dried gels [67]; a reasonable to good correspondence between experimental and theoretical values was obtained. In addition, the same concepts have been used to offer insight into how fluid and cell walls interact to define texture in vege-

tables [15,68] and how the requirement for mechanical support affects plant height [69].

A noteworthy feature of the compressive stress–strain curve for all cellular solids, regardless of which materials constitute the cell wall, is the long plateau of little or no rise in stress over a large range of strain (Fig. 5). Collapse of the cell wall once a critical stress has been exceeded is responsible for this plateau. For the same cell wall material, relative density affects the shape of the stress–strain curve, thereby affecting the mechanical properties derived from the curve. Mechanical properties are predicted on the basis of equations of a form similar to Eqs. (4) and (5), where an appropriate mechanical property of the cell wall material, a geometrical constant, and the structural parameter raised to some power define the mechanical property of interest for the cellular solid. Thus, in defining a mechanical property of relevance to crispy foods, catastrophic failure, rather than yielding, is appropriate [72]. Therefore, the failure strength of the solid material, σ_{fs} (approximately the tensile failure stress [2]), is substituted in Eq. (5) for σ_{ys}:

$$\sigma_f^* = \sigma_{fs} C_3 \left(\frac{\rho^*}{\rho_s}\right)^n \tag{6}$$

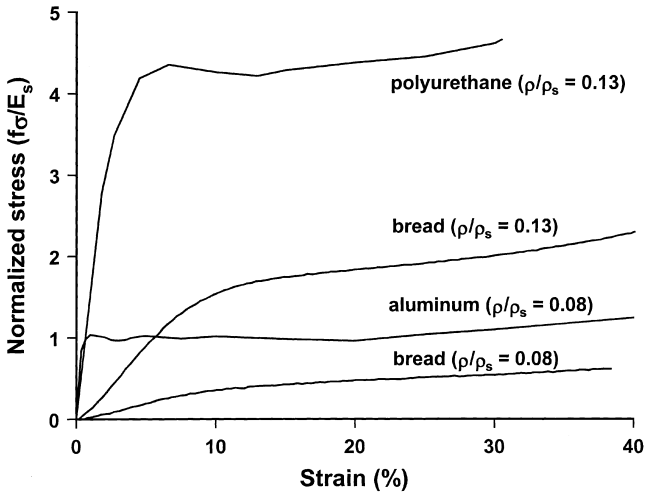


Figure 5 Normalized stress ($f\sigma/E_s$) versus strain for bread crumbs of two relative densities ($f = 500$, $E_s = 0.47$ MN/m^2), a polyurethane foam ($f = 5{,}000$, $E_s = 2.15$ GN/m^2), and an aluminum foam ($f = 50{,}000$, $E_s = 69$ GN/m^2). (Redrawn from Refs. 55, 70, and 71, respectively.)

One conundrum that biogenic cellular solids pose in the analysis of mechanical properties according to cellular solids theory, which is not dealt with by Gibson and Ashby, is the presence of water as a constituent of the solid phase. Adding more water will increase the density of the cellular solid, although it will likely decrease cell wall density, thus increasing relative density. However, the magnitude of the mechanical property of interest is unlikely to increase as a result of this increased relative density because of the effect of moisture on the properties of the cell wall [62]. Typically, moisture plasticizes biopolymers in the cell walls, lowering mechanical parameters such as E_s, σ_{fs}, and σ_{ys} [39,62]. Another issue that can confound adequate definition and prediction of the properties of biogenic cellular solids according to the theories of Gibson and Ashby is that the properties of the cell wall solids are sensitive to the processing conditions used to create the walls [13]. This issue is not unique to biogenic cellular solids [6], and some of the distribution in wall properties is attributable to the various hierarchies of structure in the solid that can be effected by processing [47]. Despite these issues, a cellular solids' treatment is instructive for the prediction of the mechanical properties of biogenic cellular solids because the response of the cell to mechanical loading reflects the response of the solid as a whole [73]. That mechanical response is, in turn, influenced by the dimensions of the cell and/or its constituent parts (Figs. 1 and 2).

1. Cell Size

Mechanical properties of cellular solids are only weakly dependent on cell size [2]. In mixed protein/maize meal extrudates that had been expanded by supercritical CO_2, cell size was not found to be a good predictor of strength, but density was [64]. On the other hand, for maize meal that had been extruded conventionally, two strength parameters of the extrudates were strongly related to a combined effect of density and cell size [74]. In the latter study, the effect of cell size on extrudate strength was obscured by other structural changes, such as changes in cell wall thickness and shape. Although cell size distribution is expected to have a greater effect on mechanical properties than cell size [2], a measure of cell size uniformity was not significantly related to the modulus and tensile strength of bread crumb [12].

2. Cell Wall Dimensions

From relations such as Eqs. (1)–(3), it is apparent that there is a direct relationship between the dimensions of the cell and its cell walls and the density of two-dimensional cellular solids such as the honeycomb. For three-dimensional structures, where cell wall formation is by random physical and/or biochemical processes, exact relationships such as Eq. (2) can only be obtained by painstaking modeling of the various structural configurations

[3,9]. Nevertheless, an exact definition of cell geometry is not necessary in order to define a scaling relationship between cell wall dimensions and relative density [2]:

$$\frac{\rho^*}{\rho_s} \propto \left(\frac{t}{l}\right)^p \tag{7}$$

The exponent, p, has a value of 2 when all cell wall material is confined to lines within the cellular solid, creating an open-cell structure, and p is 1 when cell wall material creates a closed-cell structure. Therefore, on the basis of Eqs. (4), (5), and (7), a cellular solid will be stiffer and stronger as the cell wall thickens and more compliant and weaker as the cell enlarges. Fracture properties are also a function of cell wall dimensions [2,75]. Resistance to fracture is generally to be maximized in plants and bones, and to be manipulated in foods in order to achieve desirable textures [8]. The critical stress intensity factor, K_{Ic}^*, defines when a crack of given length will propagate in a cellular solid loaded in tension provided crack length is much larger than cell size [75]:

$$K_{\mathrm{Ic}}^* = C_4 \sigma_{\mathrm{fs}} \sqrt{\pi l} \left(\frac{t}{l}\right)^2 \tag{8}$$

From Eqs. (7) and (8), it can be seen that the resistance to fracture will also scale with relative density This was seen to be so for wood [2], although fracture properties were strongly influenced by anisotropy in cellular structure; a 10-fold difference in the critical stress intensity factor was observed for cracks propagated across the grain rather than along it.

In experimental studies examining the modulus of cellular solids, the constant C_1 in Eq. (4) was found to have a value close to 1 for engineering foams [2], wheat starch bread crumb [66] and freeze-dried bread crumb [26]. Therefore, for open-cell gas-filled cellular solids [2],

$$\frac{E^*}{E_s} = \left(\frac{\rho^*}{\rho_s}\right)^2 \tag{9}$$

Despite the lack of a definitive unit cell for the cellular structure of biogenic cellular solids, the theoretical relationships of Eqs. (4) and (9) have been used to relate microstructure to modulus. Strong relationships between the modulus (from compression and flexure tests) and the relative density of various types of extrudate have been found [13,62,64]. The modulus of higher-moisture (>25%) biogenic cellular solids has also been well described by scaling relationships to relative density, including the modulus of bread in compression [41] and tension [12,55] and starch bread in compression [66]. However, researchers have reported discrepancies between experimental

values of the power-law index, m, and its theoretical value of 2 for open-cell cellular solids and 3 for closed-cell cellular solids [12,13].

Small-strain analyses have also been used to examine relationships between the elastic modulus and relative density in biogenic cellular solids. The relationship between elastic modulus and relative density is shown in Fig. 6 for freeze-dried bread crumb and fresh bread crumb determined by low-intensity ultrasound [26] and dynamic mechanical analysis (DMA) [76], respectively. For the freeze-dried bread crumb, the value of the exponent is 1.98, very close to the theoretical value of 2 [Eq. (9)], although a higher value is found by DMA for the fresh bread crumb.

From Eqs. (5)–(7), it is apparent that cell wall dimensions will affect the strength of the cellular solid as measured by a critical stress value such as σ_y^* or σ_f^*. Good relationships between σ_f^* in compression and relative density have been found for brittle extrudates [13]. For biogenic cellular solids which yield plastically, such as wood, σ_y^* is also strongly related to relative density [2]. For bread, which collapses elastically in compression [52], the theoretical value of $n = 2$ was a good fit for data on the collapse stress of starch breads [66], but when bread fails in tension, where its failure mode is plastic, values for n were lower than the theoretical value of $n = 1$ [12].

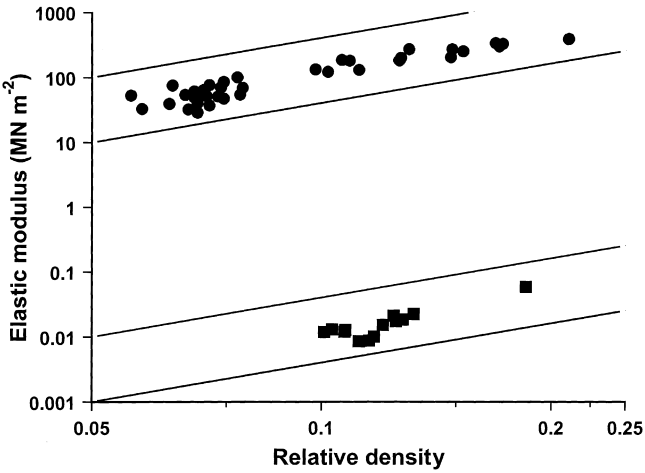


Figure 6 Relationship between elastic modulus and relative density for fresh bread crumb (■) determined by DMA [76] and for freeze-dried bread crumb (●) determined by ultrasound [26]. Lines are slopes of 2 (i.e., modulus scales with the square of the relative density).

A discussion of the significance of values for the exponents m and n in terms of cell structure and expected modes of deformation of the cell walls for brittle starch extrudates has been given by Warburton et al. [63]. The manner in which the liquid phase distributes itself in the cell walls during foam fabrication and at the onset of rigidification is expected to affect mechanical properties that determine useful properties such as texture and disposability, in an analogous manner to how the distribution of the solid within the cell has an effect on the mechanical performance of engineering foams [5].

3. Missing Cell Walls

Numerical simulations of the stress–strain behavior of periodic and non-periodic honeycombs have shown that both the elastic modulus and compressive strength [45] decrease as cell walls are removed. A decrease in modulus and strength of 30% was apparent when 5% of the cell walls was removed. Loss of 35% of the cell walls led to the complete reduction of both modulus and strength [45], a catastrophe attributable to the generation of a plane of disconnectivity in the honeycomb structure.

Although modeling the effect of missing cell walls on the elastic modulus can be reduced to analyses of the numbers of missing cell walls, as inferred from the above statements on tomographic analyses of structure, interaction between stress fields generated by two closely spaced defects will degrade strength more than the effect associated with just the loss of solid material. Therefore, strength will be markedly affected by differences in spatial distribution of these missing cell walls. Uniformity in the distribution of cell wall thicknesses of bread crumb was positively correlated with both elastic modulus and tensile fracture stress [12], whereas the effect of missing cell walls alone was not [12].

Ingredients and processing will affect the number of missing cell walls that arise during foam formation, thereby affecting the resulting properties of the product. Emulsifiers reduce the number of missing cell walls in foaming [5] and, therefore alter texture in the product. For example, in brittle extrudates, a greater number of structural ruptures (associated with cell wall failures) was observed when an emulsifier was added to a maize grits base [77]. Coalescence of cells during processing eliminates cell walls, affecting both the texture and the appearance of the product [36].

4. Anisotropy

Anisotropy in cellular solids has two direction-dependent categories: structural anisotropy and material anisotropy [2]. Structural anisotropy is a measure of cell shape, whereas material anisotropy refers to direction-dependent properties of the cell walls.

For cells with one dimension longer than the other two, structural anisotropy is characterized by a shape anisotropy ratio (R), measured as the ratio of the largest cell dimension to the smallest [2]. Two values of R are necessary to characterize structural anisotropy where cells sizes differ in all three dimensions. Effects of cell anisotropy on mechanical properties have been reported for wood [2], cork [2,65], brittle extrudates [43,63,64], and bread crumb [78].

Material anisotropy is evident for cell walls in plants, with the orientation of cellulose microfibrils creating a strong dependence of modulus on direction of testing [19].

5. Fluid Effects

Interaction of the fluid phase with the solid network of the cellular solid can lead to substantial modifications of the measured mechanical properties. For open-cell cellular solids, the stress attained at a given value of strain may increase due to movement of the fluid through the pores of the cell wall network. Three factors determine the magnitude of this enhancement of stress level: the size of the pores, the viscosity of the fluid, and the rate at which stress is applied to the cellular solid [2]. A consequence of the large difference in viscosity between air and water is that only if the pores are small will strain rate effects be observed in air-filled open-cell cellular solids. However, when the fluid is aqueous, even cellular solids with large pore sizes will exhibit fluid flow effects, which will modify their stress–strain curves at high strain rates. Such effects will have a bearing on textural quality because of the high rates of compression that are operative during mastication [8].

When cells are closed, fluid flow is not possible without breakage of the cell wall (or cell membrane in respiring biogenic cellular solids). Prior to breakage, the bulk modulus of the fluid will contribute to the stress that must be applied to produce a given strain. For a gas at atmospheric pressure, this additional contribution to the modulus is considered small [2]. However, for biogenic cellular solids, where the modulus of the solid material may only be equivalent to a few atmospheres, the gas may make a significant contribution to the value of the initial modulus (provided the walls or membranes do not fail). In this case, Poisson's ratio of the cellular solid exerts more of an effect on the modulus than relative density does. When the fluid phase is liquid, liquid incompressibility will induce stretching of the cell walls to accommodate the deflection of the cell walls that have been bent into the cell by the applied stress. This stretching of the cell walls causes all of the wall material to be "elastically active" and the exponent, m, in Eq. (4) is essentially 1 [15], particularly where the volume fraction of the solid phase is low, as is the case for plant cellular solids. The fluid therefore enhances the stiffness of biogenic cellular solids significantly.

D. Transport Properties

Various molecules or forms of energy can propagate through cellular solids. How well these energies or molecules are transported affects both end-product characteristics and processing behavior [79]. Measurements of transport phenomena can also be used as a means of probing the structure of cellular solids [80]. Propagation may occur predominantly through either the solid or the fluid phase or through the "effective medium" that is representative of the overall microstructure of the cellular solid. For instance, if the solid phase is highly conductive, then the flow of electricity is confined to the cell wall assembly; conversely, the transport of molecules is essentially restricted to the fluid phase because diffusivity in solids is considerably slower than in fluids. Microstructural details will evidently be important: Larger numbers of unconnected cell walls will decrease electrical and thermal conductivity, whereas closed cells will restrict the passage of migrating molecules. Because transport properties are a critical factor determining the suitability of biogenic cellular solids for a given application, four properties will be examined to illustrate the relationship between cellular microstructure and transport properties.

1. Transport of Molecules

Molecules migrate either in response to a gradient in some thermodynamic potential or because a stress is applied to the cellular solid which generates a pressure gradient. In either case, small molecules entrapped in the cellular solid, such as those that contribute to flavor and aroma, will migrate. This is evidently detrimental from a food quality perspective if these molecules are lost during the course of the product's shelf life or during processing, but it is beneficial if their migration contributes to a flavor sensation during mastication. For open-cell cellular solids, where a molecule is diffusing in the fluid phase and cell size is much larger than molecular size [81], diffusivity of the molecule in the cellular solid, D^*, relative to its diffusivity in the fluid alone, D, will be a function of the volume fraction of the solid phase (relative density if the fluid is a gas). An upper bound for the diffusivity when the molecule cannot penetrate the solid has been given by Kim and Torquato [81]:

$$\frac{D^*}{D} \leq \frac{1 - 1.21\varphi + 0.21\varphi^2}{1 + 0.29\varphi} \tag{10}$$

From replotting Thijssen's results [82] for the retention of a homologous series of alcohols in freeze-dried maltodextrin (Fig. 7), it can be seen that relative density does affect the retention of the migrating molecule and that an interaction between molecular size and relative density also occurs. Processing conditions, both during freezing and freeze-drying, affect the microstruc-

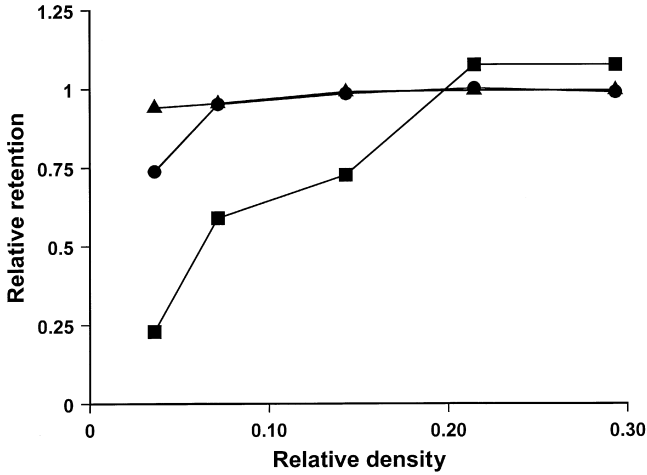


Figure 7 Relationship between retention of alcohol (relative to retention in solid maltodextrin) and relative density of freeze-dried maltodextrin. [Redrawn from data of Ref. 82 for low-pressure freeze-drying with alcohols: *n*-pentanol (▲), *n*-propanol (●), and methanol (■).]

ture of the cellular solid and hence alter retention of a given molecule over the course of the product's shelf life [82]. Similar effects can be seen in other cellular solids [30].

The transport of molecules in response to a pressure gradient has important connotations for both the texture and the flavor of biogenic cellular solids, particularly where the fluid is an aqueous solution of characteristic impact compounds. Where cells are closed, the catastrophic failure of the cell walls causes immediate release of the fluids, thereby imparting a sensation of juiciness [8].

2. Thermal Transport

In cellular solids containing a gaseous fluid phase, heat flux is poorer than in the solid materials from which they were constructed. The inherently poorer thermal conductivity of food materials exacerbates the low thermal transport capacity of biogenic cellular solids. Thermal conductivity is one component of thermal diffusivity; the volumetric heat capacity is the other; in cellular solids, both are functions of the volume fraction of solid material.

The overall thermal conductivity of a cellular solid is given by the sum of two conduction terms, one convection term and one radiation term [83], and it has been expressed as [2]:

$$k^* = \tau k_s \varphi + k_f (1 - \varphi) + k_c^* + k_r^* \tag{11}$$

where k_s and k_f are the thermal conductivity of solid and fluid, respectively, τ is a factor dependent on cell wall structure, k_c^* is the conductivity associated with convective heat transfer through the fluid, and k_r^* is conductivity associated with radiative heat transfer through the cellular solid. The convective term is highly dependent on cell size, scaling with its cube [2,79]. Although the validity of the correlations between Nusselt number and Raleigh number is questionable for cellular systems, if one takes typical maxima in cell sizes [200 μm for potato parenchyma (fluid is aqueous) and 10 mm for gas cells in croissants (fluid is air)], then the convective term k_c^* will contribute significantly to k^*. For liquid-filled cells, the radiative contribution is essentially zero due to thermal absorptivity of the liquid, but in air-filled cellular solids, radiative transfer depends strongly on the radiation absorption coefficient of the cell walls and their thickness. Therefore, as in transport of heat by convection, k_r^* is also a function of cell size for a given solid volume fraction [2]. Radiative heat transfer is also strongly dependent on temperature, and because food quality considerations limit how hot cell walls can be, k_r^* is a minor consideration in the transport of heat in biogenic cellular solids. Consequently, the first two terms in Eq. (11) dominate in evaluations of the thermal conductivity of biogenic cellular solids, whereas the magnitude of the third term's contribution depends on cell size and temperature gradients operative during processing.

Poor thermal conductivity is evidently undesirable in baking and freezing operations. In the baking of bread, where gas cells in the outer regions of the loaf are compressed due to the continued expansion of gas cells in the non-solidified interior of the loaf, the concomitant decrease in τ [Eq. (11)] retards heat flux into the center of the loaf (cf. acoustic analogy in Section IV.D.4), thus slowing rigidification of the interior portions. Conversely, poor thermal conductivity is a desirable attribute if biodegradable packaging is being fabricated from bioenic cellular solids.

The other component of thermal diffusivity is the heat capacity per unit volume given by

$$C_v^* = C_{v_s}\varphi + C_{v_f}(1-\varphi) \tag{12}$$

where C_{v_s} and C_{v_f} are the volumetric heat capacity of the solid and fluid phases, respectively. For cells filled with air, the second term on the right-hand side of Eq. (12) is small due to air's low density, and so the heat capacity of cellular solids will be reduced compared to their solid material. The converse is true when the fluid phase is liquid because water has a higher heat capacity than any biogenic solid material and any component that it is likely to dissolve [20]. Precise data on the thermal diffusivity of biogenic cellular solids are scarce; Table 2 lists some ranges in values.

Table 2 Range of Thermal Diffusivities Reported [20] for Selected Biogenic Cellular Solids (Approximately Room-Temperature Values)

Biogenic cellular solid	Thermal diffusivity (10^{-7} m^2/s)
Apple	0.8–1.5
Bread	1.7–5.6
Cake	0.9–2.1
Grapefruit	0.8–1.2
Potato	1.2–1.8

3. Electrical Properties

The ability to pass or to restrict the passage of electrical energy in biogenic cellular solids is not as important as the transport of mass and thermal energy. However, knowledge of electrical properties is relevant in some food processing operations such as ohmic heating [84], where, for example, pieces of fruit in a pie filling are heated by dissipation of electrical energy as the current propagates through the cellular solid structure. Electrical properties also appear to be important in bones where electrical or electromagnetic stimulation can promote the healing of fractures [85].

Because the current will take the path of least resistance, the electrical conductivity of a cellular solid is a function of the resistivities of the solid and fluid phases, their relative proportions, and how they are structurally organized. In cellular solids where the fluid is an aqueous solution of electrolytes, current will preferentially pass through the high-conductivity liquid phase; if cells are closed, overall conductivity will be low. For those biogenic cellular solids where membranes (rather than cell walls) enclose the fluid phase, significant passage of current only occurs when electroporation has been induced in the membranes. This precipitous reduction in resistivity is put to good use in pulsed electric field technology aimed at destroying individual bacterial cells as a result of leakage of electrolytes through their cell walls following membrane breakdown.

The low electrical conductivity of air forces the current to follow a path of low resistance through the network of the solid phase in air-filled cellular solids. Therefore, the electrical conductivity, s^*, will be some function of the relative density and the conductivity of the solid phase material, s_s [2,51]:

$$s^* = s_s z \left(\frac{\rho^*}{\rho_s}\right)^t \tag{13}$$

For highly conducting solids such as aluminum, the exponent, t, has been found to be 1.55–1.75 [51]; the same authors report z to be close to 1. Biogenic solid materials are not highly conducting, with some components, if dry and ion-free, having a resistivity comparable to that of air, and there is a 10^{21}-fold difference between the resistivities of air and aluminum. Resistivity is highly dependent on moisture content: Wood's overall electrical resistivity changing almost 10,000-fold for a 20% change in moisture content [2]; a 1000-fold change in the resistivity of dehydrated ground carrots was reported over a similar increase in moisture content [86].

An electrical property that has considerable bearing on the quality and convenience of food products is the dielectric coefficient. Direct current determinations of the dielectric coefficient of air-filled cellular solids indicate a linear relationship with relative density [2]:

$$K^* = 1.001 - \left(\frac{\rho^*}{\rho_s}\right)(K_s - 1.001) \tag{14}$$

where K_s is the dielectric coefficient of the solid phase. Data compiled for wood [2] verify this relationship, with greater slopes observed when the wood has a higher moisture content because K_s is strongly influenced by the large dielectric coefficient of water.

In the gigahertz frequency range (of relevance to the defrosting and heating of foods), the imaginary part of the dielectric coefficient, the dielectric loss factor, is pertinent because it determines how well microwave energy is attenuated and, thus, how effectively microwaves penetrate the cellular solid. The dielectric loss factor has been reported to scale linearly wih relative density for rigid PVC foams [2]. In freeze-dried potatoes, it has a very low value [87], and when moisture was used to increase relative density, an interaction between components in the solid and fluid phases was evident, with the value of the dielectric loss factor being higher at 50–80% moisture than it is in pure water [87].

4. Acoustic Permeability

The acoustic properties of some biogenic cellular solids, such as cork and wood, are put to good use in abating sound or in amplifying it [2], but for the food subset of cellular solids, these properties have not been exploited (with the exception perhaps of Rice Krispies). Nevertheless, acoustic properties can be used as a tool for probing the structure and physical properties of cellular solids. In particular, acoustic waves at ultrasonic or audible frequencies can be propagated through a cellular solid at a range of frequencies to provide a sonic spectograph of the product. One major difficulty with propagating sound through cellular solids is that the structure of the cellular solid is the source of many interfaces between the two phases. If the difference between

the acoustical impedances of the two phase is large, then sound will be strongly reflected at each cell wall. The multiple scattering that occurs may be such that it precludes transmission measurements of the sample. If this is the case, then measurements of the acoustical impedance of the cellular solid, rather than its acoustic propagation, may provide useful information on its structure [88].

If sound can be transmitted effectively through the cellular solid, two parameters can be derived from the measurements: the velocity and the attenuation coefficient. The phase velocity of acoustic waves that are polarized in the longitudinal direction (v_l) is given by

$$v_l = \sqrt{\frac{B + 4G/3}{\rho^*}} \tag{15}$$

where B and G are the bulk and shear moduli, respectively. For cellular solids comprised of a gaseous fluid, Poisson's ratio is small, so that $B + 4G/3 \sim E$. Therefore, if the Gibson and Ashby prediction for the modulus of an open-cell cellular solid as a function of relative density [Eq. (9)] is correct, then [26]

$$v_l^* = v_{ls}\sqrt{\frac{\rho^*}{\rho_s}} \tag{16}$$

where v_{ls} is the longitudinal velocity in the cell walls provided their thickness is much smaller than the wavelength of the sound. For bread that had been proofed for various time to vary crumb density and then freeze-dried ($80 < \rho^* < 280$ kg/m^3), a very good fit was observed for the square-root dependence of ultrasonic velocity (at 50 kHz) on density [26], confirming the validity of Eq. (16).

Elmehdi et al. [26] also manipulated density by compressing the bread samples prior to freeze-drying them. The anisotropy engendered in the samples led to large differences in the velocity of ultrasonic propagation, depending on whether sound was propagated perpendicular or parallel to the direction in which the sample had been strained (Fig. 8). An interpretation of this result is that the sound propagates through the cell walls, so that the slower velocity that is observed in the parallel sample is due to the longer path that the sound must take in crossing the cellular solid. Therefore, in anisotropic cellular solids, differences in structure can be probed by propagation of sound in orthogonal directions.

For biogenic cellular solids where the fluid is a liquid, the mismatch in acoustical impedances between cell walls and cells is much less. Therefore, the reduction in velocity and increase in attenuation due to scattering of the sound is not as pronounced. Higher velocities and lower attenuation coefficients have been reported for vegetables with low air contents, such as

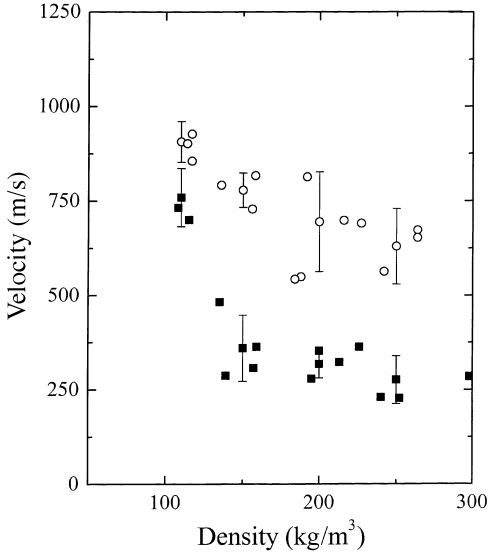


Figure 8 Ultrasonic velocity (at 50 kHz) along two orthogonal directions in a compressed then freeze-dried bread crumb: parallel (■) and perpendicular (○) to the applied compressive stress. Error bars represent standard deviation of the velocity for four ranges in density (± 25 kg/m^3). (Reprinted from Ref. 26, with permission from Elsevier.)

potatoes [89], although the state of stress that is engendered in the cellular assembly by turgor pressure has a dramatic effect on the velocity and attenuation [90]. In other liquid-filled cellular solids, even small amounts of air-filled "cells" substantially reduce the velocity and increase the attenuation [89].

V. FUTURE PROSPECTS

Fabricating structures with desirable end-use properties (including the possession of appealing textural contrasts) while minimizing the usage of materials is an attractive feature of biogenic cellular solids. In addition, the high porosity of biogenic cellular solids allows them to be potentially used as vehicles for the delivery of compounds such as pharmaceuticals and nutraceuticals to the consumer. Therefore, their use, both short and long term, will undoubtedly increase.

One area that is an active area of research and development is incorporation of agricultural materials such as starch and protein within polymer

systems to create foams with better biodegradability. The rationale for this research is not only to add extra value to agricultural commodities but also to use more environmentally friendly foaming processes [91]. The ability to fabricate cellular solids exclusively from agricultural materials will depend on how well biomaterials perform as replacements for conventional cellular solids such as polyurethane foams, factoring in the cost of the petroleum-based and the agricultural raw materials.

Another area with exciting prospects for creation of biogenic cellular solids is through the use of a commercially proven technology in the food industry—supercritical fluid carbon dioxide processing [92]. Although these microcellular structures have so far been fabricated from nonbiological molecules, the use of supercritical fluid CO_2 coupled with extrusion does permit a high density of gas cell nuclei to form in starch melts [93]. The ability to create a myriad of cellular structures by judicious manipulation of concentration of the molecule and supercritical processing conditions [92] opens up the possibility of producing a wide variety of biogenic cellular solids through the selection of appropriate molecular groups. The environmentally friendly properties of supercritical CO_2 compared to conventional production processes infers that processing will be "clean" [92], whereas supercritical CO_2's generally-regarded-as-safe production process infers that any novel biogenic cellular solids produced by this process will obviate protracted regulatory scrutiny.

One other prospect worth mentioning is the potential to create biogenic cellular solids of highly reproducible cellular dimensions through templating production processes. Kondo et al. [94] recently reported that cellulose microfibrils could be created in an ordered uniaxial alignment by constructing a template of glucan chains and using a cellulose-synthesizing bacterium. *Acetobacter xylinum*, the organism responsible for nata de coco, followed the track of the glucan chains, depositing cellulose microfibrils along the template as it progressed. Therefore, a wide variety of cellulose-based cellular solids could be constructed from appropriately configured glucan templates. Another biopolymer with the potential for creation of square honeycomb structures of precise cell size is branched DNA [95]. The stiffness of the chains and the defect-free structures resulting from their self-assembly offers the possibility of high-performance nanocellular solids where the cells can be used to contain a variety of molecules with industrial, biomedical, or consumer applications.

ACKNOWLEDGMENTS

I am grateful to NSERC (Canada) for research funding, to the individuals and companies who gave permission for inclusion of figures, to Dr. Elmehdi

of the University of Manitoba and Dr. Sun of Kansas State University for data used to prepare Fig. 6, and to Marie-Josée Barrett for helpful discussions.

REFERENCES

1. Gibson, L.J.; Ashby, M.F. *Cellular Solids: Structure & Properties*; Pergamon Press: Oxford, 1988.
2. Gibson, L.J.; Ashby, M.F. *Cellular Solids: Structure and Properties*, 2nd Ed.; Cambridge University Press: Cambridge, 1997.
3. Weaire, D.; Hutzler, S. *The Physics of Foams*; Clarendon Press: Oxford, 1999.
4. Durian, D.J.; Weitz, D.A.; Pine, D.J. Phys. Rev. A 1991, *44*, 7902–7905.
5. Simone, A.E.; Gibson, L.J. Acta Mater. 1998, *46*, 2139–2150.
6. Simone, A.E.; Gibson, L.J. Acta Mater. 1998, *46*, 3929–3935.
7. Gibiansky, L.; Torquato, S. J. Geophys. Res. 1998, *103*, 23,911–23,923.
8. Bourne, M.C. *Food Texture and Viscosity: Concept and Measurement*, 2nd Ed.; Academic Press: San Diego, CA, 2002.
9. Warren, W.E.; Kraynik, A.M. J. Appl. Mech. 1997, *64*, 787–794.
10. Roberts, A.P.; Knackstedt, M.A. J. Mater. Sci. Lett. 1995, *14*, 1357–1359.
11. Lakes, R. Nature 1993, *361*, 511–515.
12. Zghal, M.C.; Scanlon, M.G.; Sapirstein, H.D. J. Cereal. Sci. 2002, *36*, 167–176.
13. Warburton, S.C.; Donald, A.M.; Smith, A.C. J. Mater. Sci. 1990, *25*, 4001–4007.
14. Likos, C.N. Phys. Rep. 2001, *348*, 267–439.
15. Warner, M.; Thiel, B.L.; Donald, A.M. Proc. Natl. Acad. Sci. USA 2000, *97*, 1370–1375.
16. Zhu, H.X.; Melrose, J.R. J. Theor. Biol. 2003, *221*, 89–101.
17. Kimura, S.; Kondo, T. J. Plant. Res. 2002, *115*, 297–302.
18. Jarvis, M.C.; McCann, M.C. Plant Physiol. Biochem. 2000, *38*, 1–13.
19. Brett, C.; Waldron, K. *Physiology and Biochemistry of Plant Cell Walls*; Unwin Hyman: London, 1990; 4–57.
20. Rahman, S. *Food Properties Handbook*; CRC Press: Boca Raton, FL, 1995; 179–224, 339–392.
21. Zghal, M.C.; Scanlon, M.G.; Sapirstein, H.D. Cereal Chem. 1999, *76*, 734–742.
22. Jones, D.; Chinnaswamy, R.; Tan, Y.; Hanna, M. Cereal Foods World 2000, *45*, 164–168.
23. Hicsasmaz, Z.; Yazgan, Y.; Bozoglu, F.; Katnas, Z. Lebensm. Wiss. Technol. 2003, *36*, 441–450.
24. Attenburrow, G.E.; Goodband, R.M.; Taylor, L.J.; Lillford, P.J. J. Cereal Sci. 1989, *9*, 61–70.
25. Zietlow, P.K.; Ajao, D.; Helser, M.A. US Patent 6207216, 2001.
26. Elmehdi, H.M.; Page, J.H.; Scanlon, M.G. J. Cereal Sci. 2003, *38*, 33–42.
27. Tiemstra, P.J. Food Technol. 1964, *18*, 915–920.
28. Leshik, R.R.; del Pilar Cobos, M. US Patent 6117473, 2000.
29. Pritchard, M.K.; Scanlon, M.G. Can J. Plant Sci. 1997, *77*, 461–467.
30. Klamczynski, A.P.; Glenn, G.M.; Orts, W.J. Cereal Chem. 2002, *79*, 387–391.

31. Choudhury, G.S.; Gautam, A. J. Food Sci. 1999, *64*, 479–487.
32. Andersson, Y.; Hedlund, B.; Jonsson, L.; Svensson, S. Cereal Chem. 1981, *58*, 370–374.
33. Campbell, G.M.; Mougeot, E. Trends Food Sci. Technol. 1999, *10*, 283–296.
34. Baker, J.C.; Mize, M.D. Cereal Chem. 1941, *18*, 19–34.
35. Chiotellis, E.; Campbell, G.M. Trans. Inst. Chem. Eng. 2003, *C81*, 194–206.
36. van Vliet, T. In *Bubbles in Food*; Campbell, G.M., Webb, C., Pandiella, S.S., Niranjan, K., Eds.; Eagan Press: St. Paul, MN, 1999; 121–127.
37. Heidolph, B.B. Cereal Foods World 1996, *41*, 118–126.
38. Bloksma, A.H. Cereal Foods World 1990, *35*, 228–236.
39. Slade, L.; Levine, H. Crit. Rev. Food Sci. Nutr. 1991, *30*, 115–360.
40. Whitfield, R.E.; Wasley, W.L. In *Chemical Reactions of Polymers*; Fettes, E.M., Ed.; Interscience Publishers: New York, 1964; 367–500.
41. Liu, Z.; Chuah, C.S.L.; Scanlon, M.G. Acta Mater. 2003, *51*, 365–371.
42. Zhu, H.X.; Hobdell, J.R.; Windle, A.H. Acta Mater. 2000, *48*, 4893–4900.
43. Hutchinson, R.J.; Simms, I.; Smith, A.C. J. Mater. Sci. Lett. 1988, *7*, 666–668.
44. Albuquerque, J.M.; Fátima Vaz, M.; Fortes, M.A. Scripta Mater. 1999, *41*, 167–174.
45. Silva, M.J.; Gibson, L.J. Int. J. Mech. Sci. 1997, *39*, 549–563.
46. Donald, A.M. Cereal Chem. 2001, *78*, 307–314.
47. Liu, Z.; Scanlon, M.G. Trans. Inst. Chem. Eng. 2003, *C81*, 224–238.
48. Bréchet, Y.; Deschamps, A.; Verdier, M. J. Phys. IV (France) 1998, *8* (Pr4), 3–12.
49. Hepworth, D.G.; Bruce, D.M. J. Mater. Sci. 2000, *35*, 5861–5865.
50. Giménez, D.; Perfect, E.; Rawls, W.J.; Pachepsky, Y. Eng. Geol. 1997, *48*, 161–183.
51. Kováčik, J.; Simančik, F. Scripta Mater. 1998, *39*, 239–246.
52. Stokes, D.J.; Donald, A.M. J. Mater. Sci. 2000, *35*, 599–607.
53. Thiel, B.L.; Donald, A.M. Ann. Bot. 1998, *82*, 727–733.
54. Underwood, E.E. *Quantitative Stereology*; Addison-Wesley: Reading, MA, 1970; 23–47.
55. Scanlon, M.G.; Zghal, M.C. Food Res. Int. 2001, *34*, 841–864.
56. Sapirstein, H.D. In *Bubbles in Food*; Campbell, G.M., Webb, C., Pandiella, S.S., Niranjan, K., Eds.; Eagan Press: St. Paul, MN, 1999; 233–243.
57. Elliott, J.A.; Windle, A.H.; Hobdell, J.R.; Eeckhaut, G.; Oldman, R.J.; Ludwig, W.; Boller, E.; Cloetens, P.; Baruchel, J. J. Mater. Sci. 2002, *37*, 1547–1555.
58. Francis, F.J.; Clydesdale, F.M. *Food Colorimetry: Theory and Applications*; AVI Publishing: Westport, CT, 1975; 405–410.
59. Dahm, D.J.; Dahm, K.D. In *Near-Infrared Technology in the Agricultural and Food Industries*, 2nd Ed.; Williams, P., Norris, K., Eds.; American Association of Cereal Chemists: St. Paul, MN, 2001; 1–17.
60. Johnsen, S.; Widder, E.A. J. Theor. Biol. 1999, *199*, 181–198.
61. Tsuchikawa, S.; Kumada, S.; Inoue, K.; Cho, R-K. J. Am. Soc. Hort. Sci. 2002, *127*, 303–308.
62. Hutchinson, R.J.; Mantle, S.A.; Smith, A.C. J. Mater. Sci. 1989, *24*, 3249–3253.
63. Warburton, S.C.; Donald, A.M.; Smith, A.C. J. Mater. Sci. 1992, *27*, 1469–1474.

64. Gogoi, B.K.; Alavi, S.H.; Rizvi, S.S.H. Int. J. Food Prop. 2000, *3*, 37–58.
65. Fortes, M.A.; Teresa Nogueira, M. Mater. Sci. Eng. 1989, *A122*, 227–232.
66. Keetels, C.J.A.M.; van Vliet, T.; Walstra, P. J. Cereal Sci. 1996, *24*, 27–31.
67. Rassis, D.K.; Saguy, I.S.; Nussinovitch, A. J. Agric. Food Chem. 1998, *46*, 2981–2987.
68. Pang, C.H.; Scanlon, M.G. Can J. Bot. 1996, *74*, 859–869.
69. Niklas, K.J. Ann. Bot. 1993, *72*, 173–179.
70. Goods, S.H.; Neuschwanger, C.L.; Henderson, C.C.; Skala, D.M. J. Appl. Polym. Sci. 1998, *68*, 1045–1055.
71. Andrews, E.; Sanders, W.; Gibson, L.J. Mater. Sci. Eng. 1999, *A270*, 113–124.
72. Vincent, J.F.V. J. Sci. Food Agric. 1998, *78*, 162–168.
73. Christensen, R.M. Int. J. Solids Struct. 2000, *37*, 93–104.
74. Barrett, A.H.; Cardello, A.V.; Lesher, L.L.; Taub, I.A. J. Texture Stud. 1994, *25*, 77–95.
75. Choi, J.B.; Lakes, R.S. Int. J. Fract. 1996, *80*, 73–83.
76. Wang, F.C.; Sun, X.S. Cereal Chem. 2002, *79*, 108–114.
77. van Hecke, E.; Allaf, K.; Bouvier, J.M. J. Texture Stud. 1998, *29*, 617–632.
78. Hibberd, G.E.; Parker, N.S. J. Texture Stud. 1985, *16*, 97–110.
79. Gekas, V. *Transport Phenomena of Foods and Biological Materials*; CRC Press: Boca Raton, FL, 1992.
80. Paek, J.W.; Kang, B.H.; Kim, S.Y.; Hyun, J.M. Int. J. Thermophys. 2000, *21*, 453–464.
81. Kim, I.C.; Torquato, S. J. Chem. Phys. 1992, *96*, 1498–1503.
82. Thijssen, H.A.C. In *Freeze Drying and Advanced Food Technology*; Goldblith, S.A., Rey, L., Rothmayr, W.W., Eds.; Academic Press: London, 1975; 373–400.
83. Fricke, J.; Hümmer, E.; Morper, H.-J.; Scheuerpflug, P. Rev. Phys. Appl. Colloq. 1989, *24* (C4), 87–97.
84. Fryer, P.J.; Davies, L.J. In *Food Processing Operations Modeling*; Irudayaraj, J., Ed.; Marcel Dekker: New York, 2001; 225-263.
85. Oishi, M.; Onesti, S.T. Neurosurgery 2000, *47*, 1041–1056.
86. Nelson, S.O. Trans. ASAE 1973, *16*, 384–400.
87. Mudgett, R.E. In *Engineering Properties of Foods*; Rao, M.A., Rizvi, S.S.H., Eds.; Marcel Dekker: New York, 1986; 329–390.
88. Kulmyrzaev, A.; Cancelliere, C.; McClements, D.J. J. Food Eng. 2000, *46*, 235–241.
89. Self, G.K.; Povey, M.J.W.; Wainwright, H. In *Developments in Acoustics and Ultrasonics*; Povey, M.J.W., McClements, D.J., Eds.; Institute of Physics: London, 1991; 129–164.
90. Scanlon, M.G.; Page, J.H.; Toews, S.R.K. Am. J. Potato Res. 2000, *77*, 418.
91. Lin, Y.; Hsieh, F.; Huff, H.E.; Iannotti, E. Cereal Chem. 1996, *73*, 189–196.
92. Huang, Z.; Shi, C.; Enick, R.; Beckman, E. Chem. Mater. 2002, *14*, 4273–4280.
93. Alavi, S.H.; Rizvi, S.S.H.; Harriott, P. Food Res. Int. 2003, *36*, 321–330.
94. Kondo, T.; Nojiri, M.; Hishikawa, Y.; Togawa, E.; Romanovicz, D.; Brown, R.M. Jr. Proc. Natl. Acad. Sci. USA 2002, *99*, 14,008–14,013.
95. Seeman, N.C. Mater. Today 2003, *6* (1), 24–29.

14

Modeling of Formation and Rheology of Protein Particle Gels

JOOST H. J. VAN OPHEUSDEN Wageningen University, Wageningen, The Netherlands

I. INTRODUCTION

The term *aggregation* is derived from the Latin word *aggregare*, which means "forming a herd," gathering cattle into a group. By extrapolation, it is applied to the gathering, active or passive, of many similar objects. In physics and chemistry, it has the meaning of the formation of larger structures from smaller units, and it is mostly restricted to physical (reversible) bonding of somewhat larger elements (cf. colloidal particles). Polymerization is usually not considered an aggregation process because it involves chemical binding, and the same applies to vulcanization, perhaps also because there is hardly any motion involved. Coagulation, such as can occur with emulsion droplets, is no aggregation, because in the process, the identity of the primary particle is irreversible lost. Aggregation is also used for micelle formation by proteins or lipids and for block copolymers forming supramolecular structures. Finally, the state of aggregation of matter in physics is synonymous with phase: solid, fluid, gas, or plasma. That does not imply that all of the processes involved in phase transitions and phase separation can be considered as aggregation or its inverse. Actually, the term aggregation is not used for any of the fundamental transitions [e.g., solid–liquid (melting–freezing), or liquid–gas (condensation–evaporation)]. However, many of the fundamental processes that establish the dynamical phase transition are aggregation processes, for example, the nucleation process in a first-order phase transition from gas to liquid or solid.

It appears that already at the level of giving a name to a process, one is modeling that process. Here, the modeling of the dynamics of aggregation of colloidal protein particles and that of the resulting aggregates is described. The system consisting of just the primary particles is unstable and what will be the structure of the more stable aggregates formed depends on the detailed

dynamics of the process. Apparently, processes with those characteristics occur in quite a variety of systems, and, consequently, we can build more detailed models for these systems that share the same fundamental ingredients. In fact also for the above systems that are not generally considered to display aggregation phenomena, polymerization, and vulcanization, many of these fundamental ingrediens are nevertheless present, and quite similar models can be used. Actually, the most important models used to describe the dynamics of aggregation processes are borrowed from these neighboring fields.

When aggregates become very large, they may form a volume-spanning network, called a gel. Essentially, there are two main types of gel: particle and polymer gels. Polymers may bind to other polymers by nonbonded interactions, such as physical entanglements, or hydrophobic–hydrophilic interaction in block copolymers or denaturated proteins. Particles such as hard silica spheres, micelles, or globular proteins can bind through, for instance, Vanderwaals forces, or depletion interaction, and flocculate. At this point, it depends on the details of the aggregations process and the physical chemistry of the system concerned whether this clustering will lead either to space spanning gel networks, possibly transient, or separation from the suspending fluid by sedimentation or creaming. Here, we treat aggregation of (protein) particles; the gel models may well apply to polymer gels as well, but, in general, the timescales and length scales involved are very much different. There are also some more fundamental differences.

First, the general elements of the process of aggregation of identical spherical particles that might have to be taken into account in any suitable model are discussed. That depends on their possible significance during the stages of the aggregation process as a whole. Note that the concept of stages constitutes a reductive modeling procedure. It implies that at different times during the process, different descriptions might be adequate.

II. PROCESSES IN AGGREGATION AND GELATION OF PROTEIN PARTICLES

A. Diffusion

A colloidal protein particle in a suspending fluid will perform Brownian motion due to the collisions with the molecules in the suspension, which leads to translational and rotational diffusion (Fig. 1). If the colloidal particles are not much larger than the suspension molecules, between separate collisions these small particles will move ballistically (i.e., in a straight line). Of course, the same applies for larger particles, but in the time between consecutive collisions, the distance a particle has traveled is very much smaller than its

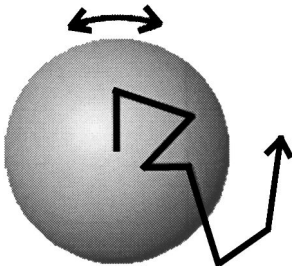


Figure 1 Rotational and translation diffusion of a single spherical protein particle in a suspending fluid.

diameter. Also for small particles, the collisions are not uncorrelated, as the number of neighboring suspension molecules is relatively small. That means that in between two collisions with the same molecule, the effect of the former collision will not have relaxed away completely. Finally, the neighbors provide a "cage" for the small colloidal particle, from which it must escape before being able to diffuse over larger distances. This caging effect is especially relevant for autodiffusion, where all particles are of the same size (e.g., in the case of a large volume fraction of colloidal particles, when collisions between colloidal particles have to be taken into account).

Clusters of particles also diffuse, but slower than individual particles (Fig. 2). How the translational and rotational diffusion coefficient scale with cluster size depends on the details of the diffusion process. Because of the larger surface area, the number of collisions per time unit with the suspension molecules will be larger, but they will, to a large degree, compensate each

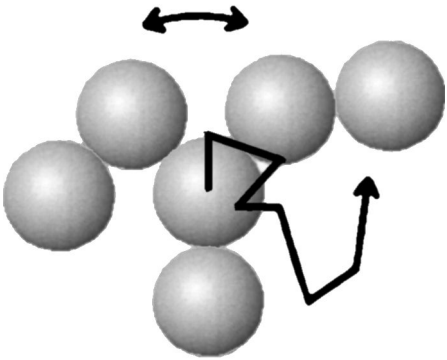


Figure 2 Translational and rotational diffusion of a cluster of particles.

other. If the collisions are fully uncorrelated, the root-mean-square (rms) average force acting upon a cluster consisting of N particles, hit by N independent forces, scales as $\sqrt{N}$. The mass of the cluster increases with a factor N, which gives a decrease in the dynamics with a factor $1/\sqrt{N}$. Similar arguments can be applied to the scaling of the rotational diffusion coefficient, but the calculation of the torques involves the structure of the cluster explicitly. For homogeneous clusters again scaling with $1/\sqrt{N}$ is found. In reality, the assumption of uncorrelated forces on the individual particles is not always valid, and other scaling relations must be used, depending on the internal structure of the clusters.

B. Hydrodynamics

Hydrodynamics describes all effects of the motion of the colloidal particles upon that of the suspension fluid. This motion then influences that of the colloidal particle itself and of other colloidal particles. A single spherical particle moving slowly through a Newtonian fluid at rest will be subjected to a Stokes friction force proportional to the diameter of the particle, the velocity of the particle relative to the fluid, and the viscosity of the suspending fluid. The only other relevant parameter is the condition at the surface, whether we have slip flow, stick flow, or something in between (Fig. 3). For non-Newtonian fluids, there are different relations.

At the level of the individual particles, only the two processes mentioned so far are relevant for the dynamical behavior of the system: that of diffusion and that of friction with the fluid. At the very early stage of aggregation, when different particles can still be considered as isolated objects, this may be the proper model. The diffusion coefficient is given by the balance between the collision forces due to thermal motion and the friction force due to particle motion. For slip flow at a low Reynolds number, we have

$$D = \frac{k_B T}{6\pi\eta a} \tag{1}$$

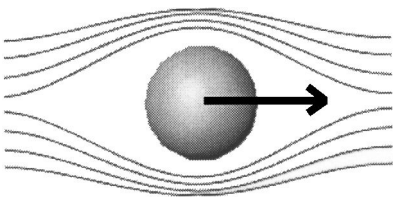


Figure 3 The motion of a sphere through a viscous fluid produces a Stokes drag force.

with k_B the Boltzmann constant, T the fluid temperature, η the fluid viscosity, and a the particle radius.

In practice, the regime over which the isolated particle picture can be maintained will be rather small. It does take some time before the Stokes flow field has developed, but when it has, it will have an infinite range over which it will influence other particles in the fluid. It is especially this long-range hydrodynamic interaction which makes problems involving hydrodynamics computationally hard. The effect of the flow field around a moving sphere is depicted in Fig. 4. The full effect of the motion of one particle upon other particles, as mediated by the suspending fluid, is much more complex than just this dipole field. Most important here is the lubrication effect. When two particle surfaces come close together, there is an additional force because of the thin fluid layer between them. If the surfaces move apart, fluid has to flow in, giving an effective attraction, and a repulsion if they move toward each other. Also, the rotation of the particles influences the flow of the medium, and through this, the rotational and translational motion of all the other particles.

C. Binding and Unbinding

One obvious prerequisite for aggregate formation is that there is a mechanism by which smaller clusters join to form larger ones. This binding mechanism can be due to attractive forces between the particles over some range, but it

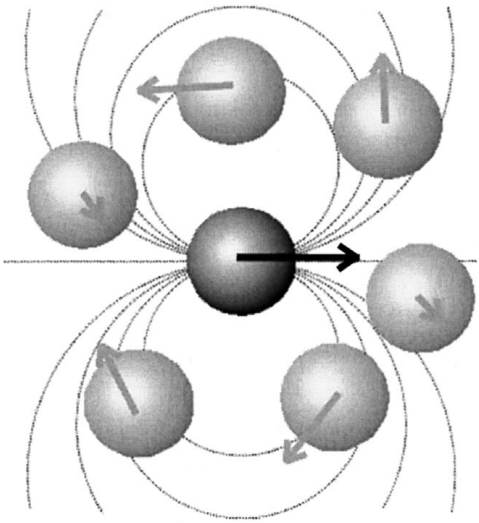


Figure 4 Effect of a moving particle on other particles in suspension.

might also be due to some reorganization taking place at the particle surfaces when they are in close contact. Coating layers of polymer material may become entangled or micelles can restructure to join with neighboring micelles. Also, vesicles and emulsion droplets can join, but they will, in general, rapidly reorganize further, leading to coalescence rather than aggregation. Micelles, emulsion droplets, and vesicles can form aggregates due to forces acting over a distance. Covalent bonding between molecules from different particles may lead to irreversible clustering.

Thermal fluctuations, both from the solvent and the particle material, may, and given enough time inevitably will, lead to breaking of these bonds (Figs. 5 and 6). The dynamic equilibrium is a balance between the internal energy gain binding and the entropy gain in unbinding.

The most important reversible interactions for destabilization and aggregation of colloidal systems are van der Waals forces, electrostatic forces, and steric, hydrophilic, and depletion interactions. Van der Waals interactions are always attractive, but, in general, rather weak. For larger particles, they may nevertheless be the most important attractive force and the cause of aggregation. Because of the r^{-6} dependence, van der Waals interactions are especially important at short range. Electrostatic forces between identical particles will always be of a repulsive nature. The presence of electric double

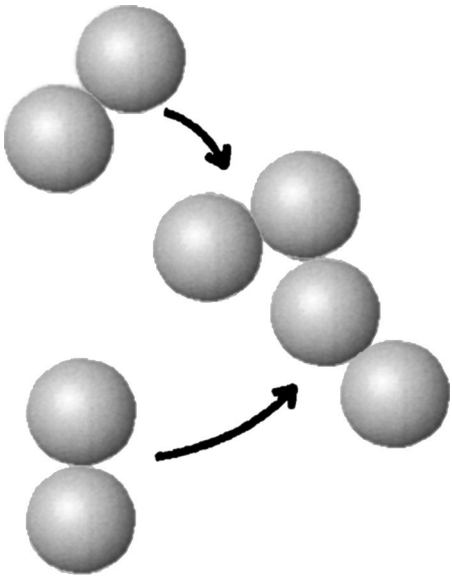


Figure 5 Cluster growth through binary particle bond formation.

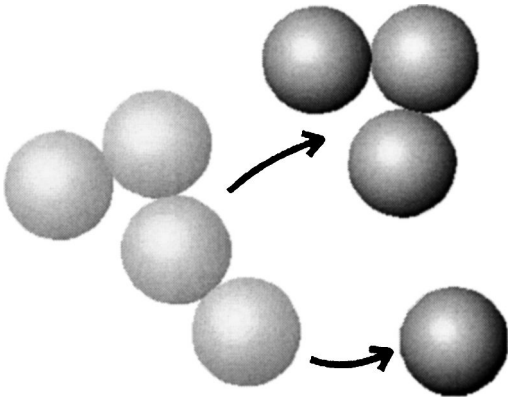


Figure 6 Cluster disruption through bond breaking.

layers, formed by the charges on the particles and the counterions in solutions, is an important source of colloid stability. Steric interactions are due to coatings of polymeric material. When the coating layers are pressed together, these molecules lose configurational freedom, and the decrease in entropy gives an effective repulsion. Steric stabilization is the main alternative to electrostatic stabilization. Hydrophilic interactions are caused by the presence of permanent electric dipoles, such as the polar headgroups of lipids and the side chains in several amino acid residues. They are important in the presence of polar solvents (e.g., when protein particles or lipid vesicles are involved) and for surfactants in emulsions. Block copolymers may be induced in forming bridges between neighboring surfaces because of these forces. Depletion interactions are attractive effective forces due to chain molecules in the solvent. Near the surface of the colloidal particles, they have less conformational freedom, which is the cause of a depletion layer of the order of the size of the globular clusters of the chain molecules in the good solvent. When two colloidal particles are close together, these depletion layers overlap, giving more freedom to the chains and hence an effective attraction of entropic origin.

The stabilization of colloids by electrostatic repulsion is successfully described by DLVO theory. In the DLVO model, there is a deep primary minimum at very short range due to the competition of van der Waals forces and steric repulsion and possibly a much more shallow secondary minimum due to the competition of the van der Waals force and the screened electrostatic repulsion. In between, there is a repulsive barrier where the electrostatic repulsion dominates over the van der Waals attraction. The existence of such a repulsive barrier before a binding between particles can take place

is probably rather common, not something exclusively reserved for DLVO systems. The effect of the barrier is that particles may come close, whereas a bond is formed only if the barrier is crossed. Also, the lubrication effect mentioned earlier can be responsible for preventing bond formation. The stability ratio W can be defined as the number of encounter, within the model, that on the average has to take place before a bond is formed. If $V(r)$ is the interaction potential, the stability ratio can be calculated using the approximation by Fuchs:

$$W = 2a \int_{2a}^{\infty} \exp\left(\frac{V}{k_B T}\right) r^{-2}\, dr \tag{2}$$

with a the particle radius. We see that both height (V) and width (dr/r^2) are important. As the model was developed for coagulation, it is expected to pertain only to two particles binding.

D. Reorganization

Binding between particles can result in cluster growth or internal reorganization. Flexibility of the strings of particles may lead to a cross-connected network of particles that may still be of a rather open structure (Fig. 7). If bonded particles roll over each other without the connection being broken, a local compactification of clusters will result (Fig. 8). Finally, bonds that break while a different bond is formed between the pieces of a cluster that have come apart upon unbinding leads to reorganization.

Reorganization need not be a constant phenomenon. For instance, with a DLVO-type binding potential with two minima, particles may at first bind loosely in the shallow secondary minimum. Thermal fluctuations easily drive the particles apart again, and the separated particles may form other bonds

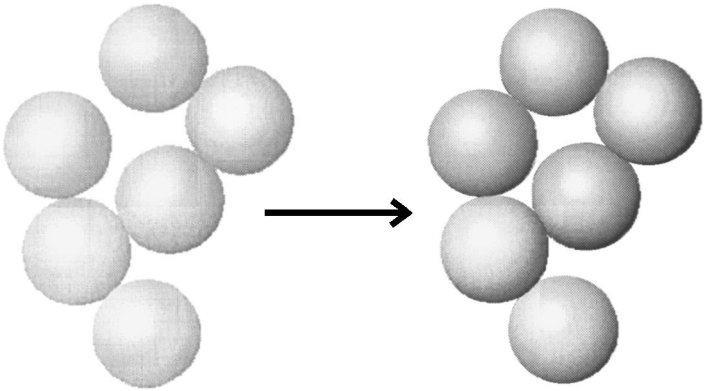


Figure 7 Bending of strings gives more connectedness of the cluster.

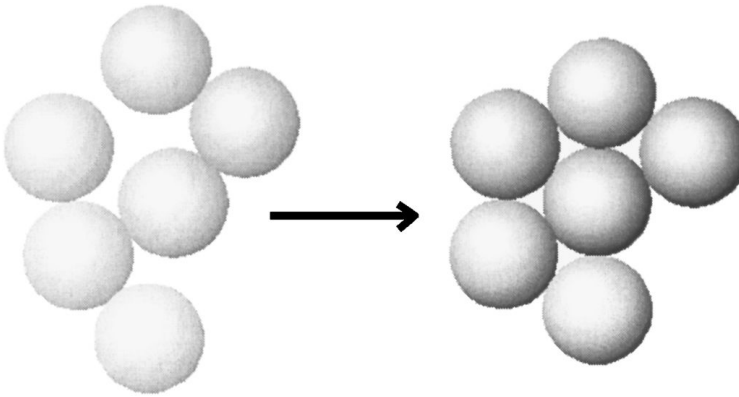


Figure 8 Rolling of particles gives local compactification of the cluster.

with particles of the same aggregate. The same thermal fluctuations that break particles apart can also urge them over the barrier into the primary minimum, from which it is much more difficult to escape. Another possibility is that the bonds themselves are not constant. A recently formed bond might be broken easily or allow for a large amount of flexibility or rolling. Subsequent development at a lower level of resolution, such as covalent bonding between molecules at the particle surface or bridging, can render bonds stiffer or less reversible.

E. Deformation and Sintering

Deformation of particles can be the result of external forces on either the individual particles or the aggregate as a whole. Particles or aggregates in flow deform due to friction forces. Stresses within the aggregate will be transferred to the particles from which it is formed and lead to bond breaking and particle deformation (Fig. 9). Especially when a gel has been formed, such stresses will become important.

Sintering of particles is due to an internal reorganization of their structure. Material moves by viscous flow or diffusion in order to reduce the surface area of the aggregates (Fig. 10). The process is driven by this interfacial energy. An isolated particle will, as a result of this mechanism, tend to maintain a spherical shape while keeping the same volume. Two such spherical particles joined together eventually fuse into a single larger spherical particle, as occurs in the coagulation of emulsion droplets, or the fusion of vesicles. Sintering is, in general, a rather slow process, so large aggregates may have formed before sintering becomes important. If sintering occurs in a larger aggregate, or even a gel, it will not always lead to macroscopic phase separation, but a bicon-

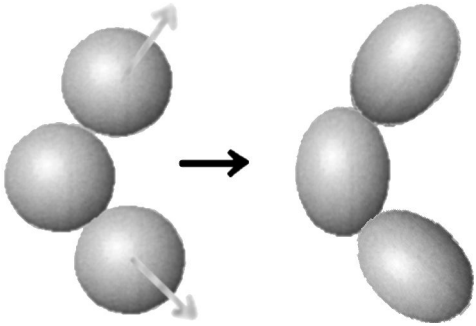


Figure 9 Stresses within a cluster may lead to particle deformation.

tinuous structure may emerge. This structure may be similar to those encountered in microemulsions, but without a vanishing of the surface tension. Instead, the remaining surface tension is balanced by the elastic bulk forces within the solid particle material.

Syneresis is a macroscopic property that certain causes gels to shrink and expel liquid. Microsyneresis is essentially the same phenomenon, that of shrinkage of the gel network, but now on a local, microscopic scale. Syneresis is therefor also called macrosyneresis. Local densification of the gel matrix can, for instance, occur in polymer networks. The total matrix can stay intact, but the pores grow. The process that drives the local compactification may not be very different from that responsible for sintering, and as such, the mechanisms are comparable. The difference is that in sintering, volume is conserved; in syneresis, it is not.

F. External Forcing

One important external force is gravity. When a colloidal suspension is destabilized, the macroscopically observable effect will, in general, be the for-

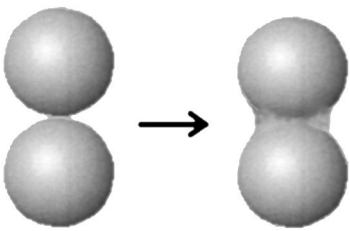


Figure 10 Sintering gives smaller, smoother surfaces.

mation of a sedimentation or creaming layer. Small colloidal particles stay in suspension because the Brownian forces are strong enough to prevent phase separation, but once they aggregate, separation takes place. The buoyancy force on a cluster simply scales with its mass, whereas the Brownian force at most scales with the square root of its mass.

Sedimentation and creaming are the result of aggregation but may also enhance it. If the aggregates are partially permeable, large clusters overtake smaller ones and bonds may be formed between them. Forced convection can be a source of aggregation, because it can act as the method by which particles or clusters are brought into contact. In the case of a simple shear flow, particles and clusters will experience different local velocities and, hence, different friction forces that tend to move them along. Still there is diffusion, which also leads to particles meeting each other (Fig. 11). When diffusion is the dominant transport phenomenon leading to cluster formation, we speak of perikinetic aggregation. If, on the other hand, internal motion is the dominant feature (e.g., when the sample is stirred), the process is called orthokinetic.

The dimensionless Péclet number (Pe) distinguishes whether diffusive or forced convection is dominant. For a single spherical particle of radius a at shear rate $\dot{\gamma}$, it is defined as

$$\mathrm{Pe} = \frac{\dot{\gamma} a^2}{2D} \tag{3}$$

with the diffusion coefficient D given by Eq. (1). For large Pe, we have orthokinetic aggregation; for small Pe, we have a perikinetic process. For particles of 1 μm diameter in water at room temperature, Pe = 1 implies shear rates of $0.04\ \mathrm{s}^{-1}$, which is very small and indicates that samples should be treated very carefully.

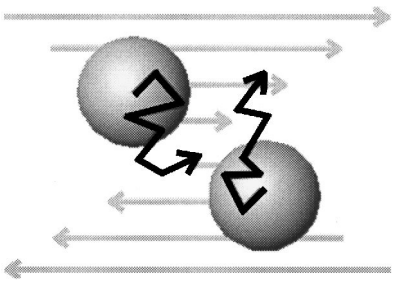


Figure 11 Competition between shear flow and diffusion.

G. Aggregation in Practice

What in practice are the relevant phenomena that influence aggregation and gelation, of course, depends very much on the system one observes. Important is that within any system, one has to be careful about the timescale at which one is looking at it. Earlier, we have mentioned many phenomena that can occur; in practice, they probably also will occur. At each given time, though, the salient behavior will be the result of a balance by only a few of these.

Destabilization of a colloidal suspension is often performed by mechanisms that can have unwanted side effects. Heating a system may result in temperature gradients that induce flows in the solvent with rates well above the ones safe to ensure perikinetic aggregation. Adding a chemical, to change the pH or salt concentration, an enzyme to change the surface properties, or polymers for depletion interaction usually involves stirring of the sample. In view of these problems, it may seem a good policy to work with particles with a high stability ratio. If aggregation is slow, there is ample opportunity to stabilize external disturbances. However, if the aggregation rate is slowed down, reorganization may become an important mechanism influencing the structure of the aggregates formed. Compactification of aggregates before they bind to others can prevent gel formation and give phase separation instead. Even without reorganization, the structures that are formed depend on the stability ratio. The higher the stability ratio, the more compact the aggregates. When the system develops into a gel, the aggregation process stops when all particles have joined the gel network. The reorganization mechanisms mentioned will lead to aging of the gel, which becomes the relevant dynamical process at longer times. All of these processes make it difficult to draw conclusions from experiments on real gels.

If we want to model aggregation and gelation, we are faced with the same problems. We cannot possibly incorporate all effects into a single model and hope to be able to perform model calculations over the full dynamical range of phenomena. Rather, we should try to develop separate models, each incorporating only those mechanisms that are relevant for the system at that give time in its evolution.

III. MODELS FOR PARTICLE AGGREGATION AND GELATION

A. The Flory–Stockmayer Model

One of the simplest models for gelation is the classical theory of Flory and Stockmayer. There is much wrong with this theory; one reason is that it does not describe the kinetics of the process. Other shortcomings will become clear

when we discuss the theory. The model deals with a polymerizing system forming bonds. As monomers, one could imagine individual mesoscopic particles. Each particle can form bonds with z other particles, and the fraction of bonds that actually has formed is given by a parameter p. In Fig. 12, is provided, an example of a structure called a Cayley tree, with light particles with z bonds depicting the core of aggregate and dark ones depicting the perimeter. At the perimeter, new bonds can be formed, but only with particles not already belonging to the cluster. We start at any particle of the cluster chosen at random and follow a path through the cluster along bonds. At each particle we reach, we have a probability $p(z-1)$ that the path can be continued. In order to have, on average, a path of infinite length, we must have

$$p(z-1) \geq 1 \rightarrow p_{\text{gel}} = \frac{1}{z-1} \tag{4}$$

When the critical fraction pgel of all bonds is reached [i.e., when, on the average, each particle has $z/(z-1)$ bonds], we have a gel. There is no explicit time dependence nor is there any spatial information involved. Each perim-

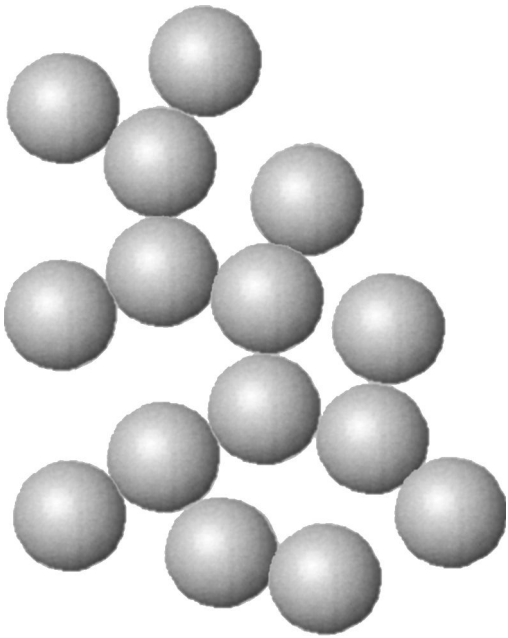


Figure 12 Example of a Cayley tree at $z = 3$.

eter particle can form a new bond. If there is enough mixing of clusters of various sizes in-between bond formation, that might be acceptable, at least at the early stages of the process. For aggregating systems close to gelation, it is very unrealistic. It can be shown that the weight fraction of clusters of size n scales as $w_n \propto n^{-3/2}$. At the gel point, the fraction of the total mass that is part of the gel hence is zero; above the gel point, this fraction becomes positive. One important feature of the model is that at all times the most common species is the monomer. At the gel point, we have a broad distribution over cluster sizes, whereas above the gel point, the relative contribution of larger clusters to the sol phase decreases with increasing p.

It can be hoped that the model has some significance if the bond parameter p is taken to increase monotonously with time. The most problematic feature is not the disability of the model to account for loops, but the absence of kinetics and lack of spatial information. Monomers have the highest diffusion coefficient, whereas the gel will not diffuse at all. Because there is only a vanishing amount of mass present in that gel, clusters attach to each other, rather than to the gel, whereas monomers would have a higher tendency to disappear than larger clusters. One disturbing feature of the neglect of rings is that the mass of a cluster of size n would scale with its radius r as $n \propto r^4$. Because the volume scales with r^D, we must have $D = 4$ in order to avoid crowding of particles within the object. This can be seen in Fig. 12. It will be difficult to add extra particles to those that still have open bonds without having the extra particles overlap.

B. Percolation Theory

In percolation theory, the paradox of the infinitely dense gel is avoided, but still there is no real kinetics. In general, the calculations within percolation theory are numerical, whereas the Flory–Stockmayer model allows analytical treatment.

Percolation models work with lattices; off-lattice calculations have the same generic results. The scaling properties of the models are independent of the lattice. Experimental validation hence should focus on these scaling results. In *site percolation* (Fig. 13), sites of a lattice are randomly filled with monomers, and neighboring monomers are assumed to be bonded. In *bond percolation*, the vertices of a lattice are randomly filled with bonds linking neighboring nodes (Fig. 14). As in the Flory–Stockmayer model, there is a critical value pgel at which an infinite cluster is formed, corresponding to the gel point. This critical value is called the percolation threshold. If the particles are considered to be spheres, with diameter equal to the node distance, for site percolation the percolation threshold can be associated with a critical volume fraction, needed to form a gel. We simply have $\varphi_{\text{gel}} = p_{\text{gel}}\varphi_0$,

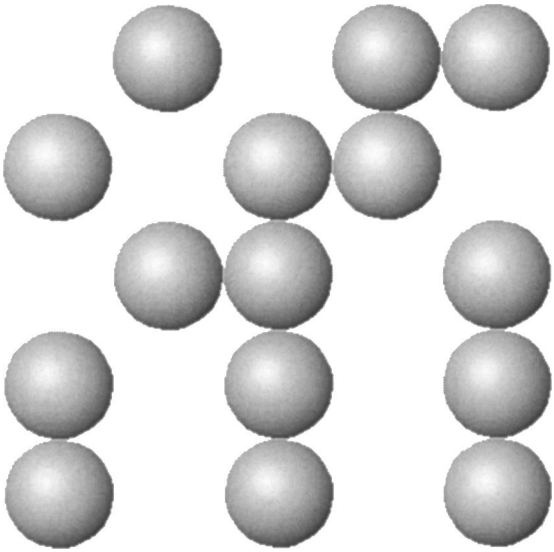


Figure 13 Square lattice in the site percolation model.

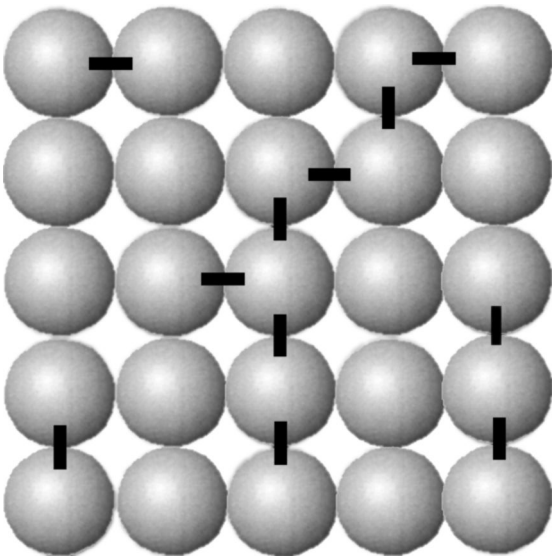


Figure 14 Bond percolation; dark particles belong to clusters.

with φ_0 the volume fraction of a single particle within the unit cell. The percolation threshold in three dimensions corresponds to $\varphi \approx 16\%$; for two-dimensional (2D) site percolation, it is 45%. These are very high values for colloidal suspensions.

Near the percolation threshold, the average cluster mass diverges. The effect is much stronger than in Flory–Stockmayer theory, where there are relatively few large clusters, even near the gel point. The gel mass fraction at the gel point is zero, but instead of increasing linearly above the critical point, it increases infinitely rapidly. This singular behavior of both the average cluster mass s and the gel fraction w is reflected by a scaling

$$s \propto (p_{\text{gel}} - p)^{-\gamma}, \qquad w \propto (p - p_{\text{gel}})^{\beta} \tag{5}$$

where β and γ are called the critical scaling exponents. For β values of 0.14 (2D) and 0.40 (3D) are found, values for γ are 2.4 and 1.7, respectively. The classical values for both variables are unity. The distribution of cluster sizes at the gel point $n_s \propto s^{-\tau}$ is very close to classical theory; in three dimensions, an exponent of 2.2 is found, instead of 2.5. The percolating cluster itself appears to be a fractal object. If one cuts out a part of the cluster of size l the mass of the selection scales with size as $s \propto l^d$, where the fractal dimensionality in three dimensions is 2.5 and $d = 1.9$ for 2D bond percolation. The same value of 2.5, and that of the other critical exponents, is also found in models for polymer gels. Whether percolation theory has any bearing to particle gels remains to be seen, as there is no kinetics involved. We will now turn to kinetic models.

C. Smoluchowski Theory

The theory of Smoluchowski was designed to describe the coagulation of emulsion droplets, but it can also be applied to irreversible particle cluster formation. The central equation of the theory is a differential equation for the number n of aggregates of mass s:

$$\frac{dn_s}{dt} = \frac{1}{2} \sum_{i+j=s} K_{ij} n_i n_j - n_s \sum_j K_{sj} n_j. \tag{6}$$

The first term gives the rate at which clusters of mass s are formed by joining two smaller clusters; the second term is the rate at which they disappear as the result of binding to other clusters of any size. The rate constants for these processes are given by the coagulation kernel K_{ij}. Only binary collisions are incorporated, indicating dilute systems.

The kinetics of the model is in the coagulation kernel. It contains both the collision rate between clusters, and in the chance of binding upon collision. For cluster formation, the latter aspect can be important. If diffusion is the main process, a power-law scaling is used:

$$K_{ii} \propto i^{\lambda} \tag{7}$$

Because the diffusion constant decreases with cluster mass, λ is expected to be negative. However, the binding probability between clusters will, in general, increase with cluster mass, so small positive values are physically meaningful. The off-diagonal elements of the coagulation kernel are often defined as

$$K_{ij} = K_{ji} \propto i^{\mu} j^{\lambda-\mu} \quad (j \geq i). \tag{8}$$

The parameter μ determines which kind of processes dominate: binding of large clusters with large clusters ($\mu > 0$) or binding of small clusters with large ones ($\mu < 0$).

Within this model, a gel is formed if $\lambda > 1$. In classical theory, any two particles can form a bond, so the reactivity of a cluster is proportional to its mass and $K_{ij} \propto ij$. That means $\mu = 0$, $\lambda = 2$, and a gel will form. For a system with $\lambda < 1$, the mean cluster mass grows with time as

$$\langle s \rangle \propto t^{\frac{1}{(1-\lambda)}} \tag{9}$$

so gelation will not occur within finite time.

In the theory of Smoluchowski, the spatial distribution of clusters affecting the kernel does not change in time. For low volume fractions, sufficient mixing of clusters in between binding could relax spatial correlation between consecutive collisions. A second drawback is that no information can be given about the detailed structure of the aggregates formed. Only the mass, not the shape, of a cluster enters the coagulation kernel.

D. Fractal Aggregation Models

In fractal aggregation models, the focus is on the internal structure of the clusters. A cluster is called fractal when the mass m within a part of that cluster of size r scales as

$$m(r) \propto r^{D_f}, \tag{10}$$

where D_f is the fractal dimensionality of the cluster. The lower the dimensionality, the more tenuous the structure. A cluster with fractal dimension near 3 is a very compact, or, rather, homogeneous, object. There are no large density fluctuations, and the mass is simply the volume times the constant density. For very small dimensionalities, near unity, the cluster consists of a

branching network of long, thin strands. As long as a single strand is contained within the sample volume, the mass will simply increase linearly with the sample size, just a bit faster when a branch is encountered. The cluster in Figure 15 has a fractal dimensionality of about 1.5.

The kinetics in fractal aggregation only models the binding of two clusters. The dynamical process of how the clusters come close to each other is not included. There are different varieties. One is the model in which individual particles bind to a cluster: particle–cluster aggregation. This could be identified with the second type ($\mu < 0$) of the Smoluchowski theory. The simulation runs as follows:

1. Place particle at the origin as seed of cluster
2. Place particle at random close to cluster perimeter
3. Randomly move particle until it hits cluster particle
4. Fix particle at present position and add to cluster
5. Goto 2.

This scheme is called Diffusion Limited Aggregation (DLA), because the rate limiting process for aggregation is the diffusion of the particles. An alternative is:

1. Place particle at the origin as seed of cluster
2. Place particle at random close to cluster perimeter

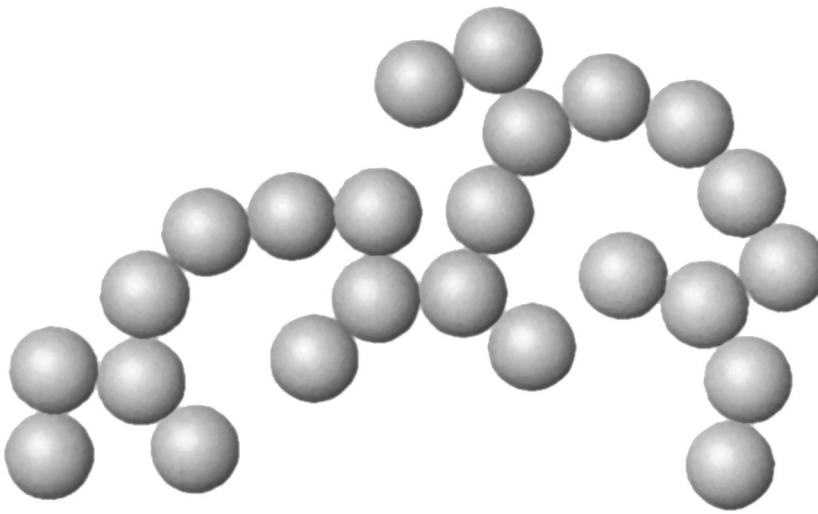


Figure 15 A fractal cluster in two dimensions.

3. Randomly move particle until it hits cluster particle
4. Select a random number x between 0 and 1
5. If ($x < \varepsilon$) Then
   6. Fix particle at present position and add to cluster
   7. Goto 2.
   Else
   6. Goto 3.
   End If

where ε is a small positive number. This process is called reaction-limited aggregation (RLA). The clusters formed in RLA are homogeneous; $D_f = 3.0$ in three dimensions against $D_f = 2.5$ for DLA. Clusters in two dimensions are $D_f = 1.7$ for DLA and $D_f = 2.0$ for RLA. In DLA, particles cannot penetrate into the core of the fractal cluster; in RLA, they can.

Simulations of cluster–cluster aggregation are more realistic. The scheme starts with a finite but large number of isolated particles, identified as clusters of unit mass:

1. Randomly select two clusters and place them close together
2. Randomly move clusters until they touch
3. Fix particles at present positions and join clusters
4. Goto 1.

This is diffusion-limited cluster aggregation (DLCA). There is also an RLCA scheme. Dimensionalities are $D_f = 1.8$ and $D_f = 2.09$ for DLCA and RLCA in three dimensions and $D_f = 1.45$ (cf. $D_f = 1.61$ in two dimensions). For cluster–cluster aggregation, reaction limitation still produces fractal aggregates. When rotation between the binding clusters is incorporated, smaller dimensionalities are found (1.3 and 1.45 for DLCA in two dimensions and three dimensions, respectively).

The above fractal dimensionalities are empirical values. One aspect that may help explain them is the so-called self-similarity of such fractal clusters. It means that when you zoom in or out on a fractal cluster, the object stays the same. On zooming in, you may see some additional features due to the enhanced resolution, but these are no different from the features that were observable at the edge of resolution before zooming in. Suppose we have two small fractal clusters of mass m_0 and size r_0. When the clusters bind, they overlap (Fig. 16). In accordance with self-similarity, we fix the overlap percentage to a value q. The radius of the new cluster becomes $r_1 = (2 - q)r_0$, and the mass $m_1 = 2m_0$. The binding of two clusters of mass $2m_0$ with overlap percentage q gives $r_2 = (2 - q)^2 r_0$, and $m^2 = 4m_0$. In general, we have

$$r_n = r_0(2 - q)^n, \qquad m_n = 2^n\, m_0 \tag{11}$$

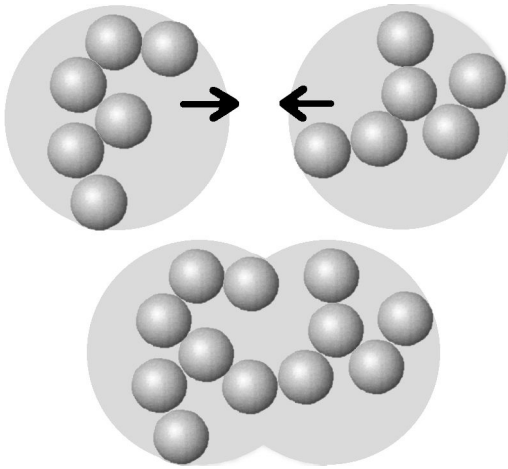


Figure 16 Two small clusters overlap when they join into a larger cluster.

from which we may derive the fractal dimensionality

$$D_f = \lim_{n\to\infty} \frac{\log(m_n)}{\log(r_n)} = \lim_{n\to\infty} \frac{\log(2^n\, m_0)}{\log(r_0(2-q)^n)} = \frac{\log(2)}{\log(2-q)} \tag{12}$$

For $q = 0$, i.e., no overlap, $D_f = 1$, so we have a linear chain. An overlap of 40% yields $D_f = 1.45$, whereas for $D_f = 1.6$, 1.8 and 2.1, we need 45%, 53%, and 60%, respectively. These are the dimensionalities found for the fractal aggregation mechanisms described earlier. The general conclusion is that the less the overlap, the lower the dimensionality. That explains why rotational diffusion gives lower dimensionalities. Apparently, all values found for the standard models follow from a rather narrow band around an overlap of 50%.

For growing fractal clusters, the mass increases with size R as R^{D_f} and the volume with R^3, so the density goes to zero. If we destabilize a colloidal suspension with a given volume fraction φ and the volume fraction inside the fractal aggregates formed is decreasing with size, at some point the process must stop. The volume fraction within an aggregate obviously cannot be smaller than that of the system as a whole. This happens for

$$\varphi_{\text{int}} = \left(\frac{R}{a}\right)^{D_f-3} = \varphi \rightarrow R = a\varphi^{\frac{1}{(D_f-3)}} \tag{13}$$

where a is the radius of the primary particle from which the cluster is built. When clusters grow beyond this size, they must overlap and form a gel.

Actually, percolation theory states that a gel is formed already for a system with a volume fraction of 16%. Clusters will form a continuous network at

$$\varphi_{\text{int}} = \left(\frac{R}{a}\right)^{D_f-3} = \frac{\varphi}{0.16} \rightarrow R = a\left(\frac{\varphi}{0.16}\right)^{\frac{1}{D_f-3}}$$
$$M = m\left(\frac{\varphi}{0.16}\right)^{\frac{D_f}{(D_f-3)}} \tag{14}$$

with m the primary particle mass. For $\varphi = 0.1$ and $D_f = 1.8$, for instance, we find $R = 1.5a$ and $M = 2m$, implying that we only need dimers to have a gel. For lower initial volume fraction, $\varphi = 0.03$, we have $R = 4a$ and $M = 12m$, and for $\varphi = 0.01$, $R = 10a$ and $M = 64m$. These figures seem too low, so a more detailed analysis is necessary; nevertheless, these equations can be used to estimate where fractal theory should go over to percolation theory.

That there is such a transition explains why in fractal aggregation it is possible to have gelation at all. If scaling of diffusion coefficients and other parameters is entered into Smoluchowski theory, fractal models appear to fall into the nongelling class. The kinetics implies that it takes infinite time to reach a cluster of infinite size. However, at some point, the assumption of binary uncorrelated collision cannot be upheld. That happens at a finite cluster size, as described earlier, and thus also at finite times. Fractal theory predicts that gelation will happen at any volume fraction, not only above a certain critical value.

E. Discussion

The Flory–Stockmayer and the percolation models focus on the existence of the gel depending on the amount of bonding, but lack kinetic information. The Smoluchowski theory describes the kinetics, but lacks the spatial information of the previous models and does not include an explicit model for the aggregate structure. Because of the binary collision approximation, it can only be valid in the early aggregation stages of dilute systems. Fractal models show a wide variety of structures, but fail when aggregating clusters cannot diffuse according to the model assumptions. Like the Smoluchowski model that would restrict them to small clusters in dilute systems, but they have been shown to be remarkably adequate also for later aggregation stages, and even gelation. Whether that is due to the fact that the main processes even close to the gel point are still of the nature assumed or to a fortuitous cancellation of up until now unincorporated effects is unclear. Hydrodynamic effects have not been implemented systematically in the models investigated, nor has rotational diffusion. As lubrication gives higher dimensionalities and rotation lower ones, neglect of both in the DLCA model could explain why, in

practice, dimensionalities of 1.8 and above are rather common, but hardly any experimental results for very low dimensionalities exist.

Many of the possibly important effects as described in the first part of this overview have not been systematically treated in models. Many real systems show a percolation threshold. If a system forms fractal aggregates, and the wide variety of models suggests that generically many will, there should not be such a threshold. However, if unbinding could occur, either through thermal motion for weak bonds or through stirring for somewhat stronger bonds, that would change things. Because larger clusters are more prone to disruption by any of these causes, growth kinetics could level off at some critical cluster size because the disruption would come into equilibrium with the formation. If that cluster size is below the size at which fractal clusters gel, there would be a critical volume fraction. Fast reorganization of fractal clusters might increase their dimensionality, hence increase the critical system size size for gel formation and also the gel time. Before that occurs, reorganization may have reduced the large aggregates to a homogeneous macroscopically phase-separated sedimentation or creaming layer. Finally, after the gel has formed, the other processes become the relevant ones and they might lead to a rapid disruption of it. Thus, perhaps a gel is formed, but it could be so weak that it would be very hard to be observed. The accuracy of present equipment and the sensitivity for measuring very low-yield stresses probably precludes this. Nonetheless, (fractal) gels could exist as transient states of the material, as, for instance, in the phase separation of simple liquids, quenched deeply below the spinodal.

IV. NUMERICAL SIMULATION TECHNIQUES

The analytical models described earlier have the advantage that generic relations can be derived, but they all lack the important aspects of real systems. Numerical simulation models can, in principle, handle any model, although one should be aware of the adage garbage in, garbage out. In general, a simulation starts with N colloidal particles in a box of size V. This can be understood as representing a sample from a macroscopic system of particle density N/V, or volume fraction $\varphi = 4\pi Na^3/3V$, where a is the particle radius. Apart from the colloidal particles, there will be a large number of solvent molecules in the sample volume. The main simplification that allows for a practical description is to neglect the discrete particle nature of the suspension fluid and only treat the parameters describing the individual colloidal particles as dynamical variables of the system. The effects of the solvent are handled through the effective interactions between the colloidal particles, such as screening, hydrodynamic interactions, and effects of the solvent on the individual particle motion, such as Brownian forces and

Stokesian friction. This approach is followed in the techniques of molecular, Brownian, and Stokesian dynamics. A different approach is taken by the lattice Boltzmann and finite-element or finite-difference techniques, where the suspending fluid is described by a discretization of a continuous medium, with the surfaces of the colloidal particles as boundaries. Again, the particulate nature of the suspension fluid is neglected, but the dynamical variables describing its flow at a mesoscopic level are taken into account. Somewhat in between is the technique of dissipative particle dynamics, which uses an averaged-out description of the suspension through particlelike objects.

A. Molecular Dynamics Technique

The simplest approach is that of molecular dynamics (MD), where the effect of the suspension fluid is almost completely absent. The time evolution of the dynamical variables describing the colloidal particle system is given by the Newtonian equations of motion

$$M\,\frac{d\boldsymbol{v}}{dt} = \boldsymbol{F}(\boldsymbol{x},\boldsymbol{v},t) = -\nabla_x V(\boldsymbol{x},\boldsymbol{v},t); \qquad \boldsymbol{v} = \frac{d\boldsymbol{x}}{dt} \tag{15}$$

with the particle positions $\boldsymbol{x}$ and velocities $\boldsymbol{v}$ as the dynamical variables. This can be easily generalized to rotating particles. The effect of the fluid can be incorporated through the potential, but only as a modification of the direct interactions between the particles. Hence, the technique is restricted to those situations where those direct interactions (e.g., due to charges or currents on the particle surface) are the dominant forces. As this is rarely the case in colloidal systems, the technique is of little use. Conceptually, the technique forms the basis on which also Brownian and Stokesian dynamics operate; that is, we only need degrees of freedom for the colloidal particles and numerically solve a set of coupled differential equations for those.

B. Brownian Dynamics Technique

More sophisticated is the approach taken by Brownian dynamics (BD), which is based on the Langevin equations of motion for the colloidal particles:

$$M\,\frac{d\boldsymbol{v}}{dt} = \boldsymbol{F}(\boldsymbol{x},\boldsymbol{v},t) - \beta\boldsymbol{v} + \boldsymbol{F}_{\mathrm{B}}(t); \;\; \boldsymbol{v} = \frac{d\boldsymbol{x}}{dt} \tag{16}$$

Here, the first term is the same potential force as in Eq. (15), the second term is the drag force which the particles experience when they move through the suspending fluid, and the third term is the Brownian force. In the simplest case, the friction coefficient has just the Stokesian form: $\beta = 6\pi\eta a$, with η the

fluid viscosity and a the particle radius. For low Reynolds numbers, generically the case, the damping forces are the dominant ones. In the large damping limit, the set of equations becomes

$$\beta \frac{d\boldsymbol{x}}{dt} = \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{v}, t) + \boldsymbol{F}_{\mathrm{B}}(t) \tag{17}$$

Some features of the fluid motion can be incorporated by giving temporal and spatial correlations to the Brownian force. Essentially, that means that one uses terms in the equations that have the same mathematical form as those of real hydrodynamics for other purposes.

C. Stokesian Dynamics

The full hydrodynamic interaction of the particles necessitates a more detailed description. In molecular dynamics, one works with point particles. In Brownian dynamics, the shape of the particles enters the equations through the drag coefficient. Hydrodynamic interactions are the result of the drag forces the moving fluid in the presence of the suspended, moving particles exerts on the surface of one individual particle. In general, the equations of Stokesian dynamics (SD) still have the form of that of Brownian dynamics, but now with much more complicated interaction terms:

$$M \frac{d\boldsymbol{v}}{dt} = \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{v}, t) + \boldsymbol{F}_{\mathrm{H}}(\boldsymbol{x}, \boldsymbol{v}) + \boldsymbol{F}_{\mathrm{B}}(t); \quad \boldsymbol{v} = \frac{d\boldsymbol{x}}{dt} \tag{18}$$

The hydrodynamic force $\boldsymbol{F}_H(\boldsymbol{x}, \boldsymbol{v})$ depends on all particle positions and velocities—in general, also the rotational velocities. It contains both the singular near-field lubrication effects between particles, as well as the slow-decaying long-range contributions of the Stokes flow field. This term also includes the drag force of a single particle, as in Brownian dynamics. Again, the inertia effects may be irrelevant over the timescales involved, in which case, the right-hand side may be set to zero. The resulting equation has the form

$$-\boldsymbol{R}(\boldsymbol{x}) \cdot \boldsymbol{v} = -\boldsymbol{R}(\boldsymbol{x}) \cdot \frac{d\boldsymbol{x}}{dt} = \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{v}, t) + \boldsymbol{F}_{\mathrm{B}}(t) \tag{19}$$

where the resistance matrix $\boldsymbol{R}(\boldsymbol{x})$, at least to first order, is a function of the particle positions and orientations only. The calculation of this resistance matrix is quite complicated, and still these equations do not explicitly take into account the motion of the suspended fluid.

D. Lattice Boltzmann Technique

In lattice gas cellular automata (LGCA), extremely simple models are used that can nevertheless simulate the Navier–Stokes equations of motion. These models consist of particles moving on a lattice with only discrete velocities. Local rules determine what happens when particles collide. In practice, the applicability of these methods is limited. The lattice Boltzmann (LB) technique, like the lattice gas automata, focuses on the suspending fluid and gives a discretized description of the continuous-flow field. In fact, it is a discretization of the Boltzmann equation that describes the flow of a real-valued distribution function f_i (in multiple component systems one for each species) defined on lattice $\boldsymbol{x}$, whereas the velocity is restricted to the lattice directions $\boldsymbol{e}_i$. The distribution function can be thought of as the number of particles at a certain lattice point flowing in the direction $\boldsymbol{e}_i$. There is a streaming step between lattice sites:

$$f_i(\boldsymbol{x} + \boldsymbol{e}_i\Delta t, t + \Delta t) - f_i(\boldsymbol{x}, t) = -\frac{1}{\tau}\left(f_i - f_i^0\right) \tag{20}$$

where τ is a relaxation time and f_i^0 is a equilibrium distribution, followed by a collision step in which the distribution functions f_i at the same lattice site are combined. The collision rules for the dynamical variables are chosen such that when the local density and momentum are averaged, the resulting macroscopic density and momentum obey the specified continuum (e.g., Navier–Stokes) equations. One way to model a suspension is to treat it as a binary fluid; a different way is to incorporate the colloidal particles by applying specific boundary conditions for the distribution functions at their surfaces.

E. Dissipative Particle Dynamics

The technique of dissipative particle dynamics (DPD) is based on the dynamical variables describing the suspended particles. The dynamical equations of motion are similar to that of MD, updating velocities in the collision step and then the positions in the propagation step:

$$\begin{aligned} \boldsymbol{v}(t + \Delta t) - \boldsymbol{v}(t) &= \frac{\Delta t}{m}\boldsymbol{F}(\boldsymbol{r}) - \gamma w_D \boldsymbol{F}_D(\boldsymbol{r}) + \sigma w_R \boldsymbol{F}_R(\boldsymbol{r}, \Delta t) \\ \boldsymbol{r}(t + \Delta t) - \boldsymbol{r}(t) &= \boldsymbol{v}(t + \Delta t)\Delta t. \end{aligned} \tag{21}$$

The right-hand side in the collision step includes the direct potential force between the particles, a friction term (between particles), which is multiplied with a weight function w_D that vanishes beyond a certain particle distance, and a random term with similar characteristics. The particles are not the colloidal particles themselves, but, rather, a lumped together description of the properties of the real colloidal particles within the range of the weight

functions. For this, the term *bags* is used. Although the appearance of the evolutionary equations of DPD is like that of MD, because of the continuous interpretations of the bags, the method is actually more close to the LB treatment, but without the lattice approximation.

F. Finite-Element/Difference Techniques

In computational fluid dynamics, various mathematical techniques have been developed to discretize the equations that describe the flow of fluids. As in LB, one may either treat the suspension as a binary fluid or use the particle surface as a boundary surface between two phases. A spatial grid x is used upon which the dynamical variables are defined. By connecting neighboring grid points, space is divided into cells, which form the basic structure.

In the finite-difference (FD) technique, one may take as dynamic variables the values of the densities and velocities on the grid points. There are many numerical schemes one may choose from, but all result in a set of coupled difference equations for the dynamical variables defined in the grid points. Alternatively, one may choose values on the cell edges, cell surfaces, or at the cell centers, or combinations of these.

In the finite-element (FE) technique, the physical variables are described by smooth interpolation functions over the cells, matched at the cell boundaries. The dynamical variables now are the interpolation coefficients. The method is most suited when the equations of motion are cast into the form of integral equations. The equations for the dynamic variables are found by applying, for instance, a least-squares fit to the integrals obtained with the approximated interpolated functions.

There are so many variants of the methods, as well as hybrid techniques like finite-volume methods, that it is not possible to write the standard type of equations for the dynamic variables that appear. Essentially, the LB technique is a special case of the FE/FD family, and one that is specifically fast numerically.

V. RESULTS OF SIMULATION TECHNIQUES

Examples of proteins or protein particles capable of forming particle gels are casein micelles, β-lactoglobulin, and soy glicinin [1,2]. Also stabilized emulsion droplets can, under some circumstances, form particle gels. Although the macroscopic rheological behavior of particle and polymer gel is very similar, the underlying microscopic mechanisms appear to be very distinct. Various model studies, both experimental and by simulation, have been performed to distinguish these mechanisms for particle gels. Here, we will discuss mainly simulation studies, and we will refer to the experiments as the "real" systems,

to which the simulations must be validated. Actual protein particle systems usually are too complex to be accessible for direct modeling, but simplified colloidal particle suspensions provide good model systems.

There is a large body of published material about the simulated rheological behavior of colloidal particle suspensions. Early reviews on the subject are by Heyes [3,4]; a very recent one is by Frenkel [5]. A recent review on milk-protein-stabilized oil in water emulsions is by Dickinson [6]. Much less is available about the behavior of particle gels, both the gelation process itself and the gel properties. One important reason is that particle gels are essentially nonequilibrium systems, whereas most of the standard statistical physical theories apply to systems in thermodynamic equilibrium. As there is much research on particle suspensions, many types of model have been developed to capture the essential features of such systems. The main two categories are continuum models and discrete (particle) models. In the continuum models, such as the FE or FD schemes used in computational fluid dynamics, the flow of material in the system is described in terms of densities defined on a discrete grid of points in space [7–11]. Also, the LB [12–14] and DPD [15,16] techniques belong to this category. Although this does not seem to be the most obvious way to describe discrete particles, some interesting results have been obtained with these techniques [17]. Still, because both the particles and the suspended fluid are modeled at the same level of detail in this approach, a substantial amount of computational resources is used for the fluid motion, which, in general, is of less interest. In the category of discrete models, the main techniques are MD and BD [3,4], and, more recently, SD [18–23]. In all cases, the numerical equations solved apply to the dynamic variables that describe the colloidal particles themselves, interacting with and through a continuous suspending fluid.

It is clear that hydrodynamic interactions are important in models for flow in colloidal suspensions. An important excuse for why they have been neglected in the earlier simulations [24–32] is that no efficient models were available at that time. Several approaches have been developed to make SD more efficient [33–35], and, consequently, there is an increasing interest in the application of the more sophisticated models [36–39]. The simpler BD models, without many-body hydrodynamics, could at least be used to investigate trends and develop further insight into the systems. For aggregation, these models may be more appropriate, provided the aggregation takes place in a system sufficiently at rest [40,41]. In real experiments, the latter requirement will not often be satisfied. Even for small colloidal particles, once they have formed larger clusters, aggregation is orthokinetic rather than perikinetic. Even for perikinetic aggregation, lubrication forces at close range during collision influence the aggregation possibility. As particles approach one another, this lubrication produces an effective repulsion. On the other

hand, when they move apart again, the flow of fluid into the widening gap produces an effective attraction. Van der Waals attraction between the surfaces of the particles will compensate for the repulsion due to the fluid flow from the gap between the particles. In practice, these effects may well cancel, and the overall result from including lubrication effects into a detailed model will be just a slowing down of the actual binding process. For an aggregation process that is diffusion dominated, that will only influence the formation of the smaller clusters, which will become somewhat more compact. Once larger clusters are formed, diffusion slows down anyhow and is always the limiting process. Later reorganization of aggregates will also lead to more compact structures and it remains to be seen whether inclusion of hydrodynamics will lead to substantially different gel structures in model calculations.

For simulations that investigate the rheological properties of the gels themselves, after they have formed, it is an open question whether the gel kinetics is strongly influenced by that of the suspending fluid. Particle gels are enthalpic, rather than entropic, as are polymer gels. That means that the kinetics of the gel is dominated by that of the bonds between the particles, whereas in a polymer gels, the dynamics of the polymer strands is the dominant feature. For local structure evolution in relatively dense gels, simple models may be well suited [42–45]. Only in very dilute gels, with relatively long strands, will the kinetics of the strands in the suspending fluid be important, but, otherwise, the kinetics of the gel is the kinetics of the connected network. A second question is how the external forcing of the gel is transferred to the bulk. At low volume fractions, the flow patterns in the suspending fluid will provide the dominating mechanism through which stresses are mediated. At a higher volume fraction, it will be the network itself that is responsible. In the latter case, the hydrodynamic effects will probably not be very important. More relevant is that in medium- and high-volume-fraction particle gel networks, when the network is mediating the external stresses, the motion of the bulk material will not be affine. For instance, an external shear deformation imposed on a sample will not lead to a linear deformation profile of all particles of the gel, as if it is behaving as a Newtonian fluid. Only the stress-carrying network strands can be expected to follow the external forcing, and any inhomogeneities in that network will result in nonaffine deformation. Hydrodynamic interactions between the individual particles in the gel will, of course, be present, and they may even be important for the actual dynamics, but it is unclear if they can be separated from the direct chemical and physical interactions that describe the bonds between the particles. In reality, external stresses will be mediated both by the network and the suspending fluid, and only a full hydrodynamic model can adequately describe that coupled motion. In large deformation rheology, when macroscopic fracture occurs, the fluid flow inside the cracks needs hydrodynamic modeling; simplified models can

only be expected to give a more or less reliable description of the earlier stages of fracture.

For illustrative purposes presented here are some results obtained with a simplified BD model neglecting all but single particle hydrodynamics, for aggregation, and small- and large-scale rheology. In practice, the large deformation behavior of gels is much more important than the small deformation behavior, which often is studied more extensively [46]. We have earlier developed a model in which colloidal particles can form flexible but irreversible bonds between their surfaces and form a network (Fig. 17) capable of reorganization [44,45]. With this model, we have studied the reorganization within particle gels and discovered that, contrary to expectation, the generic fractal structure of the networks is not due to the generic nature of the fractal aggregation model itself, but to the reorganization. Freshly formed gels at different volume fractions have different fractal dimensionalities, and only upon aging of these gels do the dimensionalities converge [47–51]. Some of these results and a large overview of more recent literature is reviewed in Ref. 52. The main new aspect is that we explicitly consider nonaffine deformation, in which the network itself is mediating the external stresses. As indicated earlier, the overall neglect of many-body hydrodynamics in our calculations renders the results useless as a prediction for actual quantitative gel properties of materials with the given specifications. The main value of these model calculations is in shedding light on the details of the mechanism underlying the macroscopic rheological behavior of the model materials investigated. Also, trends can be obtained from the results, on what the effect on the model system behavior of changing of parameters such as fluid viscosity.

In the nonaffine simulation model [53], a small-amplitude external sinusoidal shear deformation deformation is exerted at the external surface of the sample only and mediated into the bulk by the gel structure itself. This

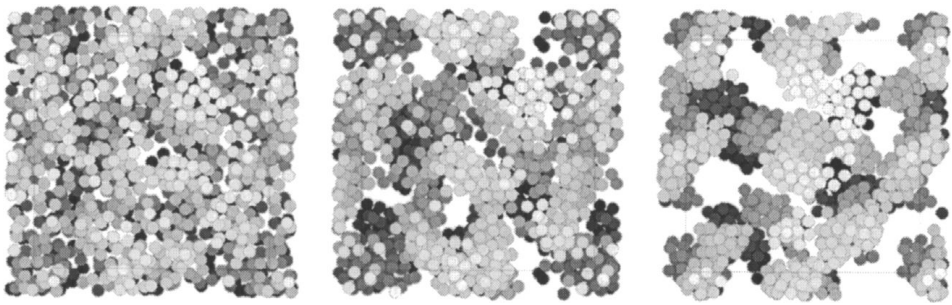


Figure 17 Evolution of a destabilized suspension to a system-spanning gel network.

approach is able to distinguish between different types of global gel structure. The more usual affine deformation, in which the gel deformation is assumed to be fluid mediated, mainly shows the local properties. Morever, the non-affine method can simulate surface load experiments. In Figs. 18 and 19, we have plotted the stress in the simulated sample for different layers. For low-frequency shear, the stress is homogeneous throughout the sample (Fig. 18); for high frequencies, the stress is high close to the surface and relaxes toward the center of the sample. There is a finite penetration depth (Fig. 19).

Large deformation shear rheology has shown a difference between polymer and particle gels. Polymer gels show a reversible transition from an elastic material to a viscous material under prolonged shear deformation. Due to the shear, the gel network breaks (Fig. 20), but it reforms at the same rate. We do not observe that behavior for our particle gel model. Increasing local stresses do lead to breaking of bonds and rupture of the network, but instead of reforming a similar global network, reorganization increases, leading to local compactification (Fig. 21). Eventually, the network is destroyed and a dense suspension of large lumpy clusters remains.

In large deformation tensile behavior, the effect of different notches was investigated, and in all cases, we find that within the parameter range investigated, ductile fracture behavior is observed. That means that if at any point of the simulation the extension of the sample is stopped, the process

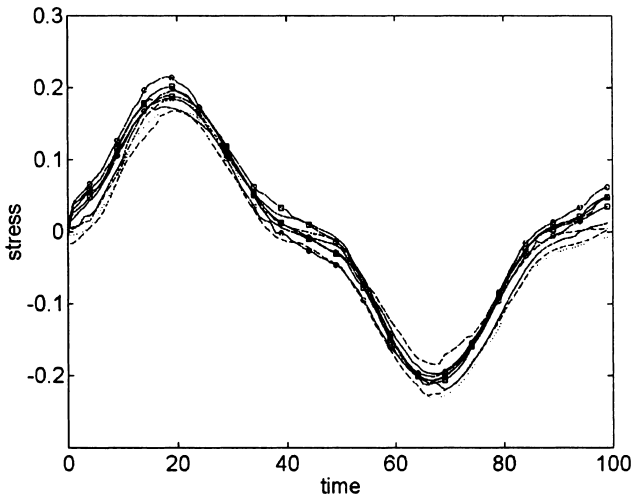


Figure 18 Stress in a periodically sheared sample at low-frequency, homogeneous stress.

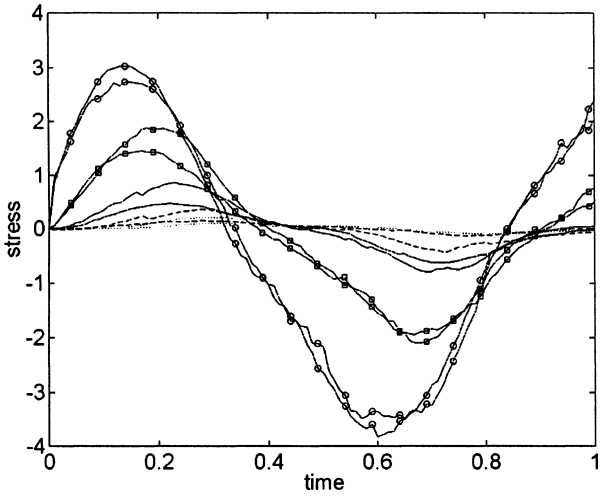


Figure 19 Stress distribution in a surface load shear experiment at high frequency. Stress is concentrated near the surface.

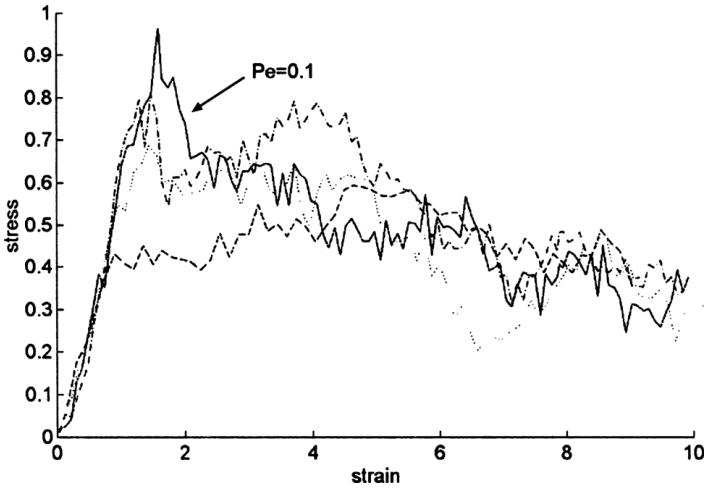


Figure 20 Stress–strain curve for large deformation shear at different shear rates.

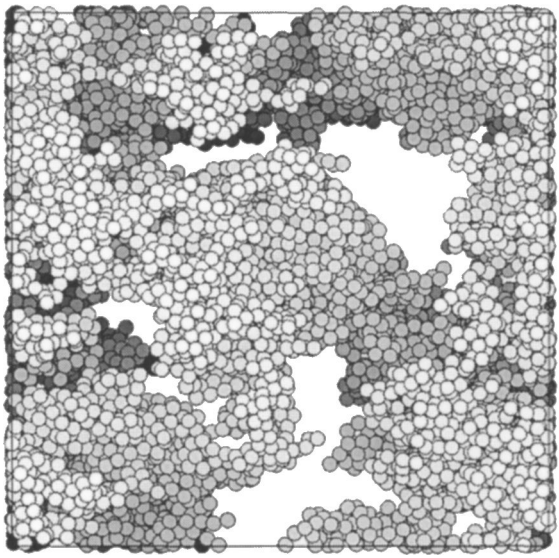


Figure 21 Lumpy structure of fractured gel network at large deformation shear.

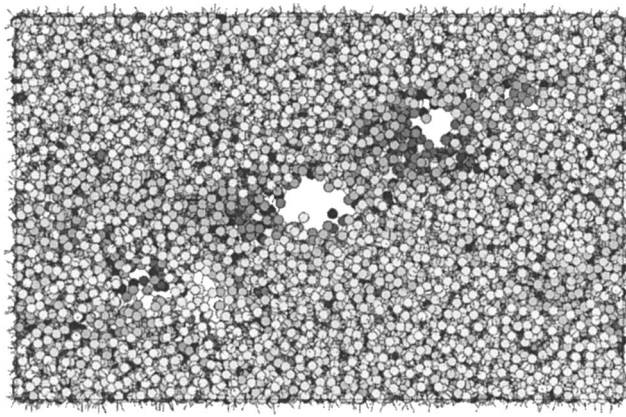


Figure 22 Fracture of a gel network under large tensile deformation in the vertical direction.

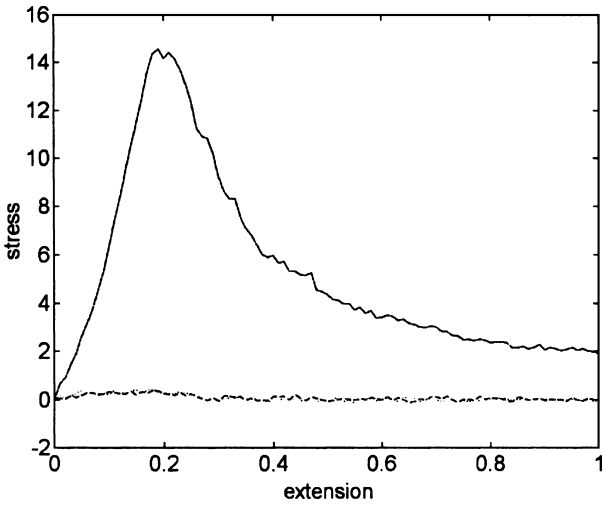


Figure 23 Stress curve for large tensile deformation in the y direction; T_{yy} (upper curve), T_{xx} and T_{zz} (lower curve).

of rupture also stops immediately. The effect of the notch is minor; in practice, rupture starts at other natural defect points in the gel matrix (Fig. 22). In the simulation, the individual components of the stress tensor can be followed (Fig. 23).

REFERENCES

1. Walstra, P.; Geurts, T.J.; Noomen, A.; Jellema, A.; van Boekel, M.A.J.S. *Dairy Technology, Principles of Milk Properties and Processes*; Marcel Dekker: New York, 1999.
2. Walstra, P. *Physical Chemistry of Foods*; Marcel Dekker: New York, 2002.
3. Heyes, D.M. J. Non-Newtonian Fluid Mech. 1988, *27*, 47–85.
4. Heyes, D.M. Adv. Colloid Interf. Sci. 1994, *51*, 247–268.
5. Frenkel, D. Physica A 2002, *313*, 1–31.
6. Dickinson, E. Colloid Surfaces B 2001, *20*, 197–210.
7. Hassonjee, Q.; Pfeffer, R.; Grantos, P. Int. J. Multiphase Flow 1992, *18*, 353.
8. Kim, S.; Karrila, S.J. *Microhydrodynamics: Principles and Selected Applications*; Butterworth-Heinemann: Stoneham, MA, 1991.
9. Lowenberg, M.; Hinch, J. J. Fluid Mech. 1996, *321*, 395.
10. Toose, E.M.; van den Ende, D.; Geurts, B.J.; Kuerten, J.G.M.; Zandbergen, P.J. J. Eng. Math 1996, *30*, 131–150.

11. Toose, E.M.; Geurts, B.J.; Kuerten, J.G.M. J. Non-Newtonian Fluid Mech. 1995, *60*, 129–154.
12. Ladd, A.J.C. J. Fluid Mech. 1994, *271*, 285.
13. Ladd, A.J.C. J. Fluid Mech. 1994, *271*, 331.
14. Ladd, A.J.C. Phys. Rev. Lett. 1996, *76*, 1392.
15. Koelman, J.M.V.A.; Hoogerbrugge, P.J. Europhys. Lett. 1992, *19*, 155.
16. Koelman, J.M.V.A.; Hoogerbrugge, P.J. Europhys. Lett. 1993, *21*, 363.
17. Boek, E.S.; Coveny, P.V.; Lekkerkerker, H.N.W.; van der Schoot, P. Phys. Rev. E 1997, *55*, 3124.
18. Bossis, G.; Brady, J.F. J. Chem. Phys. 1984, *80*, 5141.
19. Bossis, G.; Brady, J.F. J. Fluid Mech. 1985, *155*, 105.
20. Durlofsky, L.; Brady, J.F.; Bossis, G. J. Fluid Mech. 1987, *189*, 21.
21. Brady, J.F.; Phillips, R.J.; Lester, J.C.; Bossis, G. J. Fluid Mech. 1988, *195*, 257.
22. Brady, J.F.; Bossis, G. Annu. Rev. Fluid Mech. 1988, *20*, 111–157.
23. Bossis, G.; Brady, J.F. J. Chem. Phys. 1989, *91*, 1866.
24. Melrose, J.R.; Heyes, D.M. J. Colloid Interf. Sci. 1993, *157*, 227.
25. Melrose, J.R.; Heyes, D.M. J. Chem. Phys. 1993, *98*, 5873.
26. Melrose, J.R. Europhys. Lett. 1992, *29*, 51.
27. Melrose, J.R. Phys. Rev. 1991, *44*, R4789.
28. Potanin, A.A. J. Chem. Phys. 1992, *96*, 9191.
29. Potanin, A.A. J. Colloid Interf. Sci. 1993, *157*, 399.
30. Doi, M.; Chen, D. J. Chem. Phys. 1989, *90*, 5271.
31. Chen, D.; Doi, M. J. Chem. Phys. 1989, *91*, 2656.
32. van der Veer, J.M.; van Opheusden, J.H.J.; Jongschaap, R.J.J. Proceedings Third European Rheology Conference; Oliver, D.R., Ed.; Elseviers Science: Amsterdam, 1990; 489–491.
33. Sangani, A.S.; Mo, G. Phys. Fluids 1996, *8*, 1990.
34. Satoh, A.; Coverdale, G.N.; Chantrell, R.W. J. Colloid Interf. Sci. 2000, *231*, 238–246.
35. Ball, R.C.; Melrose, J.R. Physica 1997, *A247*, 444–472.
36. Jones, R.B. J. Chem. Phys. 2001, *115*, 5319.
37. Silbert, L.E.; Melrose, J.R. J. Rheol. 1999, *43*, 673.
38. Silbert, L.E.; Farr, R.S.; Melrose, J.R.; Ball, R.C. J. Chem. Phys. 1999, *111*, 4780.
39. Catherall, A.A.; Melrose, J.R. J. Rheol. 2000, *44*, 1.
40. Soga, K.G.; Melrose, J.R.; Ball, R.C. J. Chem. Phys. 1998, *108*, 6026.
41. Soga, K.G.; Melrose, J.R.; Ball, R.C. J. Chem. Phys. 1999, *110*, 2280.
42. Whittle, M.; Dickinson, E. Mol. Phys. 1997, *90*, 739–757.
43. Hütter, M. J. Colloid Sci. Interf. 2000, *231*, 337–350.
44. Bijsterbosch, B.H.; Bos, M.T.A.; Dickinson, E.; van Opheusden, J.H.J.; Walstra, P. Faraday Disccus. 1995, *101*, 51–64.
45. Bos, M.T.A.; van Opheusden, J.H.J. Phys. Rev. 1996, *E53*, 5044–5050.
46. van Vliet, T.; Walstra, P. Faraday Discus 1995, *101*, 359–370.
47. Mellema, M.; van Opheusden, J.H.J.; van Vliet, T. *Food Emulsions and Foams*; Dickinson, E., Rodriguez-Patino, J.M., Eds.; The Royal Chemical Society: Cambridge, 1999; 307–317.

48. Mellema, M.; van Opheusden, J.H.J.; van Vliet, T. J. Chem. Phys. 1999, *11*, 6129–6135.
49. Mellema, M.; Heesakkers, J.W.M.; van Opheusden, J.H.J.; van Vliet, T. Langmuir 2000, *16*, 6847–6854.
50. Mellema, M.; van Opheusden, J.H.J.; van Vliet, T. J. Rheol. 2002, *46*, 11–29.
51. Mellema, M.; Walstra, P.; van Vliet, T.; van Opheusden, J.H.J. Adv. Colloid Interf. Sci. 2002, *98*, 25–50.
52. Dickinson, E. J. Colloid Interf. Sci. 2000, *225*, 2–15.
53. Rzepiela, A.A. Deformation and fracture behaviour of simulated particle gels. Thesis, Wageningen University, 2003.