KINETICS AND THERMODYNAMICS OF MULTISTEP NUCLEATION AND SELF-ASSEMBLY IN NANOSCALE MATERIALS

ADVANCES IN CHEMICAL PHYSICS
VOLUME 151

KINETICS AND THERMODYNAMICS OF MULTISTEP NUCLEATION AND SELF-ASSEMBLY IN NANOSCALE MATERIALS

ADVANCES IN CHEMICAL PHYSICS
VOLUME 151

Edited by

GREGOIRE NICOLIS AND DOMINIQUE MAES

Series Editors
STUART A. RICE

Department of Chemistry
and
The James Franck Institute, The University of Chicago,
Chicago, Illinois

AARON R. DINNER

Department of Chemistry
and
The James Franck Institute, The University of Chicago,
Chicago, Illinois

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Catalog Number: 58-9935

ISBN: 978-1-118-16783-0

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTRIBUTORS

V. DEGIORGIO, Dipartimento di Elettronica, Università di Pavia, 27100 Pavia, Italy

ANTHONY D. DINSMORE, Department of Physics, University of Massachusetts, Amherst, MA 01003, USA

JUAN MANUEL GARCÍA-RUIZ, Laboratorio de Estudios Cristalográficos, IACT (CSIC-UGR), Avda. del Conocimiento s/n, P.T. Ciencias de la Salud, E-18100 Armilla, Spain

M. GIGLIO, Dipartimento di Fisica, Università di Milano, 20133 Milano, Italy

PATRICK GROSFILS, Center for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, CP 231, Blvd. du Triomphe, 1050 Brussels, Belgium

MATTHIAS KELLERMEIER, Institute of Physical and Theoretical Chemistry, University of Regensburg, Universitätsstrasse 31, D-93040 Regensburg, Germany; Physical Chemistry, University of Konstanz, Universitätsstrasse 10, D-78457 Konstanz, Germany

JOHN J. KOZAK, DePaul University, 243 South Wabash Avenue, Chicago, IL 60604-2301, USA

WERNER KUNZ, Institute of Physical and Theoretical Chemistry, University of Regensburg, Universitätsstrasse 31, D-93040 Regensburg, Germany

JAMES F. LUTSKO, Physics Department, Center for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, CP 231, Blvd. du Triomphe, 1050 Brussels, Belgium

DOMINIQUE MAES, Structural Biology Brussels, Vrije Universiteit Brussel, Brussel, Belgium

EMILIO MELERO-GARCÍA, Laboratorio de Estudios Cristalográficos, IACT (CSIC-UGR), Avda. del Conocimiento s/n, P.T. Ciencias de la Salud, E-18100 Armilla, Spain

CATHERINE NICOLIS, Royal Meteorological Institute of Belgium, 3 Avenue Circulaire, 1180 Brussels, Belgium

GREGOIRE NICOLIS, Interdisciplinary Center for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, Campus Plaine, CP 231, Blvd. du Triomphe, 1050 Brussels, Belgium

LIQUAN PEI, Department of Physics, University of Massachusetts, Amherst, MA 01003, USA

M.A.C. POTENZA, Dipartimento di Fisica, Università di Milano, 20133 Milano, Italy; Dipartimento di Elettronica, Università di Pavia, 27100 Pavia, Italy

T. SANVITO, Dipartimento di Fisica, Università di Milano, 20133 Milano, Italy

JOHN R. SAVAGE, Department of Physics, University of Massachusetts, Amherst, MA 01003, USA; Department of Physics, Cornell University, Ithaca, NY 14853, USA

MIKE SLEUTEL, Structural Biology Brussels, Vrije Universiteit Brussel, Brussel, Belgium

ALEXANDER VAN DRIESSCHE, Laboratorio de Estudios Cristalográficos, University of Granada, Granada, Spain

TITUS S. VAN ERP, Center for Surface Chemistry and Catalysis, K.U. Leuven, Kasteelpark Arenberg 23, B-3001 Leuven, Belgium

PETER G. VEKILOV, Department of Chemical and Biomolecular Engineering and Department of Chemistry, University of Houston, Houston TX 77204-4004, USA

PREFACE

There is increasing awareness that in many nanoscale systems the formation of ordered phases involves nonstandard nucleation and growth processes. Extensive experimental, theoretical, and computational efforts have been reported in a variety of cases, from simple fluids to protein solutions and colloidal matter. They reveal among others the occurrence of metastable phases, multistep nucleation, intermediate-mediated self-assembly and roughening types of transitions, interfering significantly with the yield of the process and the quality of the resulting material.

Despite the potential interest of this emerging field there is so far no systematic attempt to bring together representative contributions of the researchers involved, summarizing the state of the art and providing ideas on future directions. To fill this gap a 2-day workshop entitled *Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanosize Materials* was organized in Brussels on March 25 and 26, 2010 with the support of the Belgian Federal Science Policy Office, the European Space Agency, the Université Libre de Bruxelles, and the Vrije Universiteit Brussel. The present volume contains contributions based to a large extent on invited talks presented in this workshop.

The chapters by Nicolis and Nicolis, Van Erp, and Potenza et al. review some recent theoretical, simulation, and experimental methodologies for analyzing and monitoring nucleation and growth-related processes. The following chapters by Vekilov, Savage et al., Lutsko and Grosfils are devoted to experimental and theoretical aspects of two-step nucleation in a variety of materials, from proteins to colloids to simple fluids. Finally, the chapters by Kozak, Sleutel et al., and Garcia-Ruiz deal with nonstandard features underlying certain crystal growth and self-assembly problems. These contributions highlight the need to reassess our views of phenomena long thought to be "classical" and well understood where, unexpectedly, nonlinear dynamics, irreversible thermodynamics, and self-organization turn out to play a prominent role. In addition to its intrinsic interest this realization should constitute a jumping-off point for applications in a number of problems of concern, where nanoscale systems are ubiquitous.

We are much indebted to Professor Rice for accepting to include this volume in the Advances in Chemical Physics series and for his constructive comments.

Gregoire Nicolis and Dominique Maes
Brussels, May 2011

PREFACE TO THE SERIES

Advances in science often involve initial development of individual specialized fields of study within traditional disciplines followed by broadening and overlap, or even merging, of those specialized fields, leading to a blurring of the lines between traditional disciplines. The pace of that blurring has accelerated in the last few decades, and much of the important and exciting research carried out today seeks to synthesize elements from different fields of knowledge. Examples of such research areas include biophysics and studies of nanostructured materials. As the study of the forces that govern the structure and dynamics of molecular systems, chemical physics encompasses these and many other emerging research directions. Unfortunately, the flood of scientific literature has been accompanied by losses in the shared vocabulary and approaches of the traditional disciplines, and there is much pressure from scientific journals to be ever more concise in the descriptions of studies, to the point that much valuable experience, if recorded at all, is hidden in supplements and dissipated with time. These trends in science and publishing make this series, *Advances in Chemical Physics*, a much needed resource.

The *Advances in Chemical Physics* is devoted to helping the reader obtain general information about a wide variety of topics in chemical physics, a field that we interpret very broadly. Our intent is to have experts present comprehensive analyses of subjects of interest and to encourage the expression of individual points of view. We hope that this approach to the presentation of an overview of a subject will both stimulate new research and serve as a personalized learning text for beginners in a field.

STUART A. RICE
AARON R. DINNER

KINETICS AND THERMODYNAMICS OF FLUCTUATION-INDUCED TRANSITIONS IN MULTISTABLE SYSTEMS

GREGOIRE NICOLIS[1] AND CATHERINE NICOLIS[2]

[1]*Interdisciplinary Center for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, Campus Plaine, CP 231 Blvd. du Triomphe, 1050 Brussels, Belgium*
[2]*Institut Royal Météorologique de Belgique, 3 Avenue Circulaire, 1180 Brussels, Belgium*

CONTENTS

Advances in Chemical Physics, Volume 151: Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanoscale Materials, First Edition. Edited by Gregoire Nicolis and Dominique Maes.

I. INTRODUCTION

Many physical systems obey, at the macroscopic level of description, evolution equations of the form[1]

$$\frac{d\bar{x}_i}{dt} = v_i(\{\bar{x}_j\}, \lambda) \tag{1}$$

Here $\{\bar{x}_i\}, i = 1, \cdots, n$ are state variables such as density, temperature, or concentration; $\{v_i\}$ represent the evolution laws, leading eventually an initial state $\{\bar{x}_i(0)\}$ toward an invariant part of phase space referred to as the attractor; and λ is a set of parameters expressing the coupling between the system and its environment.

In most situations of interest the v_i's are nonlinear functions of the $\{\bar{x}_j\}$'s. One of the principal signatures of such nonlinearities is the multiplicity of simultaneously available states that can be reached in the long time limit. On the other hand, at the level of description afforded by Eq. (1) a transition between these states, signaling a qualitative change of behavior of the underlying system, can only be realized by the action of an external preparation bringing the initial condition from the domain of attraction of the initially prevailing asymptotic state to that of the new one. If these two regimes are separated by a finite distance the corresponding initial perturbation needs to be quite massive, and hence hardly realizable under normal conditions.

Now systems composed of several subunits—as is typically the case of physical systems—are subjected to spontaneous deviations from the state predicted by a description like in Eq. (1) owing to microscopic-level processes such as thermal motion and interparticle collisions. Furthermore, they are as a rule embedded in a complex environment with which they continuously exchange matter, momentum and energy. To account for these universal sources of variability we augment Eqs. (1) by the addition of random forces $\{R_i(t)\}$. This leads to a set of stochastic evolution equations of the form [1–3]

$$\frac{dx_i}{dt} = v_i(\{x_j\}, \lambda) + R_i(t) \tag{2}$$

We refer to the descriptions afforded by Eqs. (2) and (1) as the mesoscopic and the mean-field description, respectively.

At the level of the observables, the action of $R_i(t)$ will be manifested in the form of fluctuations around the deterministic path as provided by Eq. (1). Our main objective in this chapter is to identify the principal features of these fluctuations starting from a set of assumptions on the nature of the random forces. A most important point for our purpose is that fluctuations provide a natural mechanism of transitions between states and much of the analysis will aim at determining the mean rates and other probabilistic properties of these transitions.

The general formulation is laid down in Section II. The case where the mean-field dynamics derives from a kinetic potential is considered in Section III, where the conditions of existence of such a potential and its connection with the thermodynamic potentials are also analyzed. We show that there exists a general class of phenomena for which this connection can be implemented and which includes a variety of nucleation phenomena associated to phase transitions, one of the principal themes of this volume. Section IV is devoted to the type of kinetic potentials compatible with two-step nucleation as observed, in particular, in protein solutions. State diagrams and phase space portraits are derived and different generic scenarios are identified. A discrete state model of transitions between states compatible with this information is derived in Section V, where the mean transition times are computed and compared with the result of full-scale stochastic simulations. It is shown that the presence of intermediate metastable states may be responsible of an enhancement in the transition rates. The repercussions of different transition scenarios on the dissipation generated by the system are studied in Section VI. The main conclusions are summarized in Section VII.

II. FORMULATION

Our starting point is Eq. (2) under the additional assumption that the random forces R_i define a multivariate stationary Gaussian white noise process,

$$\left\langle R_i(t)R_j(t')\right\rangle = 2\epsilon D_{ij}\delta(t-t') \tag{3}$$

it being understood that D_{ij} form a positive definite matrix and are time independent (stationary process). ϵ is a small parameter related, depending on the case, to the thermal energy $k_{\mathrm{B}}T$ or to the inverse of the system size.

As is well known, under the above conditions the stochastic variables $\{x_i\}$ undergo a diffusion process in phase space, whose probability density P satisfies the Fokker–Planck equation [1–3]

$$\frac{\partial P}{\partial t} = -\sum_{i=1}^{n}\frac{\partial}{\partial x_i}v_i(\{x_j\})P + \epsilon\sum_{i,j=1}^{n}\frac{\partial}{\partial x_i}D_{ij}\frac{\partial P}{\partial x_j} \tag{4}$$

There is an abundant literature on both the stationary and time-dependent solutions of Eq. (4) in the case of a single variable, but the situation becomes markedly more involved in the presence of two or more variables. We summarize hereafter some results of special relevance for the purposes of the present chapter.

A. Stationary Distribution

Let us write Eq. (4) in the form

$$\frac{\partial P}{\partial t} = -\sum_i \frac{\partial J_i}{\partial x_i} \tag{5a}$$

where the ith component J_i of the probability flux $\mathbf{J}$ is given by

$$J_i = v_i(\{x_j\})P - \epsilon \sum_j D_{ij} \frac{\partial P}{\partial x_j} \tag{5b}$$

For P to be stationary the divergence of $\mathbf{J}$ must vanish. Now, in a multivariate system subjected to nonequilibrium constraints this does not imply automatically the vanishing of $\mathbf{J}$ itself, since a circulating probability flow with nonzero curl may subsist, even if the state is globally stationary [4]. This entails that the stationary solution of (4) cannot be found analytically by a simple algorithm. Nevertheless, the presence of an ϵ factor in Eq. (4) suggests to seek approximate stationary solutions of the form [1–3]

$$P_s(\{x_j\}) = \exp[-\Phi(\{x_j\})/\epsilon] \tag{6}$$

Substituting into (4) and keeping dominant terms (in ϵ^{-1}) one obtains a Hamilton–Jacobi type equation satisfied by Φ [5, 6],

$$\sum_i v_i \frac{\partial \Phi}{\partial x_i} + \sum_{ij} D_{ij} \frac{\partial \Phi}{\partial x_i} \frac{\partial \Phi}{\partial x_j} = 0 \tag{7}$$

which can be solved approximately by expanding Φ in an appropriate basis.

There are two cases where the above difficulties do not arise: One variable systems; and closed physical systems, where P_s must coincide with one of the classical distributions given by equilibrium statistical mechanics. This latter type of systems is of special relevance in the problem of multistep nucleation and will be considered in detail in Section III.

An important point is that Φ possesses a local extremum along the solutions of the mean-field equations (1). On these grounds it has been suggested that it may be viewed as a generalized entropy-like function [7]. This analogy will be taken up again in Section III.

B. Exit from an Attraction Basin

Consider now the case where the system admits several simultaneously stable steady-state solutions $\bar{x}_{i,\alpha}$, $\alpha = 1, \cdots, M$ and let C_α be the corresponding attraction basins. Clearly, C_α partition the full phase space into nonoverlapping cells

separated by boundaries constituted by the stable manifolds of the unstable steady-state solutions of Eqs. (1) $\bar{x}_{i,\beta}$, $\beta = 1, \cdots, N$. Within each C_α the vector field $\{v_i\}$ will tend toward $\{\bar{x}_{i,\alpha}\}$. Still, under the effect of a noise defined on an unbounded support, as in Eq. (3), a trajectory emanating from a point inside cell C_α will sooner or later cross the boundaries of this cell and will find itself in the domain of attraction of another stable fixed point. Two natural questions arise then in this context.

(i) Let $\mathbf{c}$ be a point on C_α. What is the distribution $\pi(\mathbf{x}, \mathbf{c})$ of points $\mathbf{c}$ by which the trajectory escapes from C_α starting from point $\mathbf{x} \in C_\alpha$?

(ii) What is the mean exit time from C_α?

Just like for the stationary distribution, an exact analytic answer to these questions is not available in the most general case of a multivariate system subjected to nonequilibrium constraints. But in the limit $\epsilon \to 0$ some asymptotic estimates can again be obtained [2]. First, the distribution of exit points is given essentially by $\exp(-\Phi(\mathbf{c}/\epsilon))$, that is, by the stationary distribution [Eq. (6)]. And second, the mean exit time is given by the inverse of the lowest eigenvalue λ_1 of the Fokker–Planck operator, which in the case of multistable systems is exponentially small, $\lambda_1 \approx \exp(-K/\epsilon)$ where K is a finite positive number determined by the structure of velocity field $\mathbf{v}$ and by the diffusion matrix D.

III. CLOSED SYSTEMS AND DETAILED BALANCE

We now limit ourselves to closed systems operating in the linear range of irreversible processes. On the one side, this implies the absence of systematic externally imposed nonequilibrium constraints, the only source of nonequilibrium being the initial deviation of $\{\bar{x}_i\}$ from the equilibrium values $\{\bar{x}_{i,\mathrm{e}}\}$. And on the other side, it implies that the fluxes $\mathbf{j}$ associated to the various irreversible processes present are connected to the associated thermodynamic forces by the relation [8, 9]

$$\mathbf{j} = L \cdot \mathbf{X} \tag{8}$$

where the matrix L of kinetic (Onsager) coefficients is assumed to be positive definite. The forces $\mathbf{X}$ are in turn related to the derivatives of a thermodynamic potential such as the free energy function F

$$\mathbf{X} = -\frac{\partial F}{\partial \mathbf{x}} \tag{9}$$

Furthermore, under an appropriate choice of variables the fluxes $\mathbf{j}$ can be identified to the time derivatives of the $\mathbf{x}$'s, it being understood from now on that these

variables belong to the class of even variables under time reversal [2, 3]. Equation (2) takes now the form

$$\frac{d\mathbf{x}}{dt} = -L \cdot \frac{\partial F}{\partial \mathbf{x}} + \mathbf{R}(t) \tag{10}$$

expressing the evolution of the state variables as the overall result of thermodynamic driven processes accounted for by F and of kinetically driven ones accounted for by L and $\mathbf{R}$. Nonlinearity and in particular the multiplicity of steady states stems in this setting entirely from F, the states being thus related to different phases in which the system of interest can exist. As for L and $\mathbf{R}$ they account for the relaxation processes and for the fluctuations around each state as well as for transitions between the states. It has to be emphasized that in writing Eq. (10) we have lumped together the spatial degrees of freedom. In this view $\{x_i\}$ represent averaged quantities, such as for instance the mass fractions of the material in different states/phases at a given stage of the transformation.

The Fokker–Planck equation [Eq. (4)] associated to Eq. (10) reads

$$\frac{\partial P}{\partial t} = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left(\sum_{j=1}^{n} L_{ij} \frac{\partial F}{\partial x_j} P + \epsilon \sum_{j=1}^{n} D_{ij} \frac{\partial P}{\partial x_j} \right) \tag{11}$$

To secure consistency with equilibrium statistical mechanics this equation must admit as stationary solution the equilibrium distribution

$$P_{\mathrm{s}} = P_{\mathrm{e}} \approx \exp(-F/\epsilon) \tag{12}$$

Inserting into Eq. (11) one sees that the diffusion matrix D and the kinetic matrix L must be proportional (a fluctuation–dissipation type relationship) and Eq. (11) simplifies further to

$$\frac{\partial P}{\partial t} = \sum_{i,j=1}^{n} \frac{\partial}{\partial x_i} L_{ij} \left(\frac{\partial F}{\partial x_j} P + \epsilon \frac{\partial P}{\partial x_j} \right) \tag{13}$$

Notice that this equation in conjunction with (12) implies a vanishing probability flux in the stationary (here equilibrium) state, $J_{i,\mathrm{e}} = 0$ for all $\{x_j\}$'s within each of the attraction basins of the fixed points. This latter property implies by itself that the velocity field must be a linear combination of the derivatives of some potential function with respect to the state variables but is actually more general than detailed balance, in the sense that it may in principle be compatible with the presence of nonequilibrium constraints.

To extract from Eq. (13) information concerning the time-dependent properties we introduce a linear transformation of variables

$$\mathbf{x} = A \cdot \mathbf{z} \tag{14a}$$

$$\frac{\partial}{\partial \mathbf{x}} = \widetilde{A}^{-1} \cdot \frac{\partial}{\partial \mathbf{z}} \tag{14b}$$

$\widetilde{A}$ being the transposed of matrix A, and set

$$F(\mathbf{x}(\mathbf{z})) = \mathrm{U}(\mathbf{z}) \tag{14c}$$

$$P(\mathbf{x}(\mathbf{z})) = \rho(\mathbf{z}) \tag{14d}$$

We assume from now on that the kinetic matrix L is state independent. By virtue of the fluctuation–dissipation theorem this applies to the correlation matrix D of the fluctuations as well (additive noise). Substituting into the equation we obtain, after some straightforward manipulations,

$$\frac{\partial \rho}{\partial t} = \left(\widetilde{A}^{-1} \cdot \frac{\partial}{\partial \mathbf{z}}\right) \cdot \left\{ L \cdot \left[\left(\widetilde{A}^{-1} \frac{\partial U}{\partial \mathbf{z}} \rho + \epsilon \widetilde{A}^{-1} \frac{\partial \rho}{\partial \mathbf{z}} \right) \right] \right\}$$

or, in more explicit form,

$$\frac{\partial \rho}{\partial t} = \sum_{jm} \left(A^{-1} L \widetilde{A}^{-1} \right)_{jm} \left(\frac{\partial}{\partial z_j} \frac{\partial U}{\partial z_m} \rho + \epsilon \frac{\partial^2 \rho}{\partial z_j \partial z_m} \right) \tag{15}$$

Equation (15) features the matrix

$$\Lambda = A^{-1} L \widetilde{A}^{-1} \tag{16}$$

linked to L by a congruent transformation. As well known under such a transformation a symmetric matrix can be diagonalized and actually, upon a further linear scaling, be reduced to the unit matrix. This requirement determines fully the transformation matrix A and upon substituting into Eq. (15), one obtains [10]

$$\frac{\partial \rho}{\partial t} = \sum_{j=1}^{n} \frac{\partial}{\partial z_j} \left(\frac{\partial U}{\partial z_j} \rho + \epsilon \frac{\partial \rho}{\partial z_j} \right) \tag{17}$$

where t is a rescaled time. This is the standard form of the Fokker–Planck equation of a system whose deterministic part of the evolution is driven by a *kinetic potential* U and which is subjected to additive fluctuations of identical variance. A system of this kind undergoes a process of isotropic diffusion in phase space described by

the Langevin equation [cf. also Eq. (10)]

$$\frac{dz_i}{dt} = -\frac{\partial U}{\partial z_i} + r_i(t) \tag{18a}$$

where the random forces satisfy the properties

$$\langle r_i(t)\rangle = 0, \qquad \left\langle r_i(t) r_j(t')\right\rangle = 2\epsilon\delta_{ij}^{\mathrm{kr}}\delta(t-t') \tag{18b}$$

As a corollary the invariant (equilibrium) probability density is given by

$$\rho_{\mathrm{s}}(\mathbf{z}) = \rho_{\mathrm{e}} \approx \exp(-U(\mathbf{z})/\epsilon) \tag{18c}$$

Equations (18) establish the possibility to cast the evolution equations of a multivariate system, under the assumptions of detailed balance and of additive noise, in a form deriving from a kinetic potential generating both the dynamical evolution and the invariant probability of the relevant variables z_i. These properties of the kinetic potential U are to be contrasted from those of quantity Φ introduced in Section II, for a system subjected to nonequilibrium constraints and not satisfying the detailed balance property. The extent to which they subsist if the matrices L and D are state dependent will be discussed briefly in the subsequent sections.

It should be pointed out that the conjunction of fluctuations and nonlinearities implies that the evolution of the mean values generated by Eq. (17),

$$\frac{\langle dz_i\rangle}{dt} = -\left\langle \frac{\partial U}{\partial z_i}\right\rangle$$

is not closed with respect to $\langle z_i\rangle$, but involves higher moments of the probability density as well. It is only when the mean-field approximation can be justified that these latter equations reduce to a closed set of deterministic evolution laws,

$$\frac{d\overline{z}_i}{dt} = -\frac{\partial U}{\partial \overline{z}_i} \tag{19}$$

Otherwise, one needs to resort to the full Eq. (18a).

Let $\{\overline{z}_{i,\alpha}\}$ be the steady-state solutions to which the solutions $\{\overline{x}_{i,\alpha}\}$ introduced in Section II transform through Eq. (14a). It follows from (19) that they are extrema of the kinetic potential U and, by virtue of Eqs. (12) and (14c), of the free energy and thus of the invariant probability itself as well. Linearizing around these states, setting $\overline{z}_i = \overline{z}_{i,\alpha} + \delta z_{i,\alpha}$ yields

$$\begin{aligned}\frac{d\delta z_{i,\alpha}}{dt} &= -\sum_j \left(\frac{\partial^2 U}{\partial \overline{z}_i \partial \overline{z}_j}\right)_\alpha \delta z_{j,\alpha} \\ &\equiv -\sum_j H_{ij}\delta z_{j,\alpha}\end{aligned} \tag{20}$$

where $H = \{H_{ij}\}$ is the Hessian matrix of U. If all eigenvalues of H are positive the reference state $\{\overline{z}_{i,\alpha}\}$ is asymptotically stable and, at the same time, a minimum of both U and F. If at least one eigenvalue of H is negative the reference state is unstable. Typically it corresponds to a saddle of the potential surface U as a function of $\overline{z}_i$ and, exceptionally, (if all eigenvalues are negative) to a maximum.

Under the conditions of existence of a kinetic potential considered in the present section the exit problem formulated in Section IIB simplifies considerably. Specifically, it can be shown that in the limit, again, of small ϵ the transition between stable states $\{\overline{x}_{i,\alpha}\}$ and $\{\overline{x}_{i,\alpha'}\}$ will follow a path across the saddle point $\{\overline{x}_{i,\beta}\}$ lying on the manifold separating the corresponding attraction basins. Furthermore, in the same limit the mean exit time depends principally on the following elements [11]:

(i) The value of the *potential barrier* ΔU, that is, the difference of values of U on the unstable (transition) state and the stable state. Since U has a local minimum on the stable state, ΔU is necessarily a positive quantity. Furthermore, by virtue of Eq. (14c) ΔU is equal in magnitude to the free energy barrier separating the states concerned by the transition.

(ii) The Hessian determinant, that is, the product of the eigenvalues of the Hessian matrix $\{H_{ij}\}$, evaluated at the minimum of the potential on the stable state.

(iii) The curvature of U on the unstable state in the direction across the saddle point.

In contrast to ΔU which as pointed out in (i) is determined from the thermodynamics, the two latter quantities depend on kinetic effects and are thus different from those associated to the free energy. Explicit expressions will be provided in Section IV for generic potentials U involving two variables.

We close this section by an application of the above outlined formalism to first-order phase transitions mediated by intermediate metastable phases [12–15]. As alluded already in the Introduction and discussed extensively in the chapters by Vekilov, Dinsmore et al., and Lutsko in this volume transitions of this kind are known to occur in a variety of materials, from protein solutions to aerosols to plasma crystals. Of special relevance is protein crystallization, where the weakness and short-range character of the attractive part of the interactions favors the existence of a long-living metastable phase in the form of a high concentration liquid. This phase tends to enhance significantly, under certain conditions, the rate of nucleation of crystals. Here we focus on the kinetic aspects of the process and, in particular, on the role of the kinetic potential U and its relationship with the free energy F. For this purpose nucleation is formulated as the dynamics of two coupled order parameters x_1 and x_2 related respectively to density—or concentration—and structure (crystallinity), evolving in an effective force field provided by the first

part of Eq. (10) and subjected to thermal fluctuations in the form of Gaussian white noise,

$$\frac{dx_1}{dt} = -L_1\frac{\partial F}{\partial x_1} - L\frac{\partial F}{\partial x_2} + R_1(t)$$
$$\frac{dx_2}{dt} = -L\frac{\partial F}{\partial x_1} - L_2\frac{\partial F}{\partial x_2} + R_2(t) \tag{21}$$

where the matrix L of the kinetic coefficients has been taken to be symmetrical by virtue of Onsager's reciprocity relations. We stress that in the absence of the intermediate phase Eq. (21) collapse into a single equation for a unique order parameter, which can be cast straightforwardly in a variational form.

Switching to the Fokker–Planck description and assuming that the fluctuation–dissipation theorem is valid we seek for a congruent transformation (14) diagonalizing L with a matrix A of the form $A_{ii} = a_i$, $A_{ij} = A_{ji} = a$. Upon carrying out the algebra one obtains

$$a_1 = \frac{L_1 \pm \sqrt{L_1L_2 - L^2}}{L}a$$
$$a_2 = \frac{L_2 \pm \sqrt{L_1L_2 - L^2}}{L}a \tag{22a}$$

and the Fokker–Planck equation in the z_1, z_2 variables

$$\frac{\partial\rho}{\partial t} = \frac{\sqrt{L_1L_2 - L^2}L^2}{a^2(L_1 + L_2 + 2\sqrt{L_1L_2 - L^2})}$$
$$\left\{\frac{\partial}{\partial z_1}\frac{\partial U}{\partial z_1}\rho + \frac{\partial}{\partial z_2}\frac{\partial U}{\partial z_2}\rho + \epsilon\left(\frac{\partial^2\rho}{\partial z_1^2} + \frac{\partial^2\rho}{\partial z_2^2}\right)\right\} \tag{22b}$$

Notice that the argument in the square root is positive, owing to the positive definiteness of the matrix L. This relation can be further reduced to the form of Eq. (17) by rescaling the variables z_1, z_2 or, more straightforwardly, the time,

$$t = \tau\frac{a^2(L_1 + L_2 + 2\sqrt{L_1L_2 - L^2})}{L^2\sqrt{L_1L_2 - L^2}} \tag{22c}$$

yielding

$$\frac{\partial\rho}{\partial\tau} = \frac{\partial}{\partial z_1}\frac{\partial U}{\partial z_1}\rho + \frac{\partial}{\partial z_2}\frac{\partial U}{\partial z_2}\rho + \epsilon\left(\frac{\partial^2\rho}{\partial z_1^2} + \frac{\partial^2\rho}{\partial z_2^2}\right) \tag{23}$$

In the context of the two-step nucleation the free energy F possesses three minima. Now, from Eq. (14) one has

$$\frac{\partial F}{\partial x_i} = \sum_{k=1}^{2} \widetilde{A}_{ik}^{-1} \frac{\partial U}{\partial z_k}$$
$$\frac{\partial^2 F}{\partial x_i \partial x_j} = \sum_{k,\ell=1}^{2} \widetilde{A}_{ik}^{-1} A_{\ell j}^{-1} \frac{\partial^2 U}{\partial z_k \partial z_\ell} \tag{24}$$

The first of these relations implies that the extrema of F transform into extrema of U, since the matrix A is nonsingular. The second relation implies that the Hessians of U and F (in their respective variables) are related by the congruent transformation used to reduce L to the unit matrix which, by construction, preserves positivity. The nature of the extrema of F is thus not affected in the evolution in terms of the transformed variables. In short, the nucleation of crystals of the solid phase is reflected by fluctuation-induced transitions removing the system, initially located in a minimum of U associated to the fluid phase, toward a minimum associated to the solid phase.

An interesting question from the standpoint of irreversible thermodynamics is whether the processes associated to the density and crystallinity (structure) fields can be thermodynamically coupled, as implied in Eq. (21) by the presence of the off-diagonal element L of the Onsager matrix. Now density is a true scalar whereas structure is accounted for by a tensor. At first sight, in an isotropic medium, this would rule out a thermodynamic coupling on the grounds of the Curie symmetry principle [8, 9]. On the other hand, a tensor T can be split in the following way

$$T = \frac{1}{3}\, I\, tr\, T + T^{(a)} + T^{(s)} \tag{25}$$

where I is the unit tensor, $tr\, T$ the trace of T that is a scalar, $T^{(a)}$ its antisymmetric part that can alternatively be viewed as an axial ("pseudo") vector, and $T^{(s)}$ its symmetric traceless part that can be alternatively be viewed as a polar ("true") vector. Clearly, then, the density field x_1 can legitimately be coupled to the first part of the decomposition in Eq. (25) or, alternatively, to any other structure related quantity x_2 of scalar nature such as the first coefficient of a Fourier series expansion of the solid-phase density. The terms in L at the level of Eq. (21) stand, precisely, for this type of coupling. Notice that even in the absence of a nondiagonal term L in this equation x_1 and x_2 would still be coupled kinetically, through the x_1, x_2 dependence of the potential.

IV. GENERIC POTENTIALS AND THEIR UNFOLDINGS

In this section, we develop some generic models of transitions between states compatible with different types of nucleation scenarios observed in nanosize materials

such as protein solutions, under the conditions of existence of a kinetic potential put forward in the preceding section. The latter is viewed as a Landau type polynomial [16, 17] that depends nonlinearly on appropriate linear combinations of the original variables—the order parameters—the main question being to determine for each given scenario the minimal number of relevant order parameters and the minimal relevant nonlinearities displayed in this potential providing a qualitative explanation of the process of interest.

Now the above described program is reminiscent of the philosophy underlying bifurcation theory [18, 19] or more appropriately catastrophe theory [20–22], which deals more specifically with the classification of the different behaviors of dynamical systems deriving from a potential. Applications of catastrophe theory to phase transitions have been the subject of several investigations [23–25], but to our knowledge they have so far been concerned with the structure of the phase diagrams of a material as some key parameters are varied. Here we also focus on dynamical aspects, including the kinetics of the transitions between different states. The particular class of phenomena that we address are transitions between a reference state, denoted by 1, to a final state, denoted by 3 it being understood that under certain conditions another state 2 may also exist such that transitions from 1 to 2 and from 2 to 3 are possible. In the literature of multistep nucleation [12–15], 1 can be a dilute protein solution phase, 3 a crystalline phase with high protein concentration, and 2 a dense protein solution phase, but in actual fact the approach is generic: it extends beyond this example and covers a wide class of materials, including colloidal systems. All states 1, 2, 3 are supposed to be locally stable and thus to correspond to minima of the kinetic potential U. We place ourselves in conditions where 3 is much more stable than 1 and 2, entailing that $U(3)$ is significantly less than $U(1)$ and $U(2)$. The relative stability of states 1 and 2 can change. Of special interest will be situations where state 2 represents a metastable phase.

The simplest "reference" situation is when state 2 does not exist. The transition between 1 and 3, a first-order phase transition, can then be understood qualitatively in terms of a Landau polynomial of 4th degree involving a single order parameter z related, for example, to the protein concentration,

$$U_4(z) = \frac{z^4}{4} + \lambda\frac{z^2}{2} + uz \tag{26}$$

This type of nonlinearity along with the two control parameters λ and u are sufficient to account for all different behaviors. Criticalities separating these behaviors are associated with the cusp catastrophe [21], one of the seven known elementary catastrophes. The transition per se from 1 to 3 occurs through an intermediate unstable state (13), lying between 1 and 2.

Let us now place ourselves under conditions that state 2 also exists. Different situations can be envisaged.

A. Transitions from 1 to 3 Occur Necessarily Through State 2

This scenario is in turn compatible with the presence of both a single or two order parameters.

(i) In the first case the minimal Landau polynomial compatible with the existence of three local minima is of 6th degree [21].

$$U_6(z) = \frac{z^6}{6} + w\frac{z^4}{4} + c\frac{z^3}{3} + \lambda\frac{z^2}{2} + uz \tag{27}$$

It involves the four control parameters w, c, λ, and u [21] and generates at criticality the butterfly catastrophe [21]. The transition from 1 to 3 involves two steps [26]. One from 1 to 2 via an intermediate unstable state (12); and one from 2 to 3 via an intermediate unstable state (23).

(ii) In the second more realistic case, in addition to order parameter z_1 related to, for example, concentration in the example of protein crystallization one also accounts for a second order parameter z_2 associated to the structure of the solid phase. We require that the associate potential switches, as the control parameters are varied, from a two-well geometry (minima at states 1 and 3) to a three-well one (minima at states 1, 2, and 3). This transition scenario is typical of the parabolic umbilic catastrophe [21] and its universal features can be captured by the potential involving four control parameters [21]

$$U(z_1, z_2) = \frac{z_1^2 z_2}{2} + \frac{z_2^4}{4} + \lambda\frac{z_1^2}{2} + \mu\frac{z_2^2}{2} + uz_1 + vz_2 \tag{28}$$

In spite of the presence of the additional variable z_2 the transition from 1 to 3 still occurs necessarily via state 2 [27] and involves two intermediate unstable states, (12) and (23) behaving as saddle points in the two-dimensional phase space spanned by z_1 and z_2.

B. Transitions from 1 to 3 Need Not Occur Through State 2

In this more flexible scenario, which is compatible with the experimental data on crystallization of protein solutions or of colloidal systems, we are led to require that all three states 1, 2, and 3 must be able to communicate with each other directly. In a two-dimensional phase space this entails three stable nodes whose attraction basins are separated by the stable manifolds of three saddle points. If only these six fixed points were present this would lead to a forbidden configuration in the form of a closed loop delimited by the unstable manifolds of the three saddles. One thus needs to stipulate the existence of a seventh fixed point, which for topological reasons must be an unstable node and thus a maximum of the kinetic potential.

To accommodate this configuration it is necessary to go beyond the potentials associated to the seven elementary catastrophes, the minimal solution satisfying the necessary genericity properties being the unfolding of the double cusp catastrophe [28, 29], which we write here in the form

$$U(z_1, z_2) = \frac{z_1^4 + z_2^4}{4} + k\frac{z_1^2 z_2^2}{2} + a\frac{z_1^3}{3} + b\frac{z_1^2 z_2}{2} + \lambda\frac{z_1^2}{2} + \mu\frac{z_2^2}{2} + cz_1 z_2 + uz_1 + vz_2 \tag{29}$$

with $k > -2$ and $k \neq 2$. This potential involves eight control parameters. It reduces along $z_1 = 0$ and along $z_2 = 0$ to a potential of the form (26) associated to the classical cusp catastrophe—whence the "double cusp catastrophe" denomination. In what follows we will be interested in the dynamical behaviors and the bifurcations generated by the partial unfolding in which we set $k = b = 0$.

The mean-field evolution equations of the two order parameters generated by the potential in Eq. (29) are [cf. Eq. (19)]

$$\frac{d\bar{z}_1}{dt} = -\bar{z}_1^3 - a\bar{z}_1^2 - \lambda\bar{z}_1 - c\bar{z}_2 - u$$
$$\frac{d\bar{z}_2}{dt} = -\bar{z}_2^3 - \mu\bar{z}_2 - c\bar{z}_1 - v \tag{30}$$

In the limit where the coefficient c is zero these equations are uncoupled. Their steady-state solutions are thus solutions of two independent cubic equations. Depending on the sign of the corresponding discriminant each of these can have one to three real solutions, the transition between these different regimes being conditioned by the cusp catastrophe. Globally therefore, barring nongeneric situations where two real solutions merge into a single one (discriminant of one of the cubics vanishes) the full system can have for given parameter values one, three, or nine real solutions.

Consider now the case of c being nonzero. Equation (30) become then coupled and their steady state solutions z_{1s}, z_{2s} satisfy the relations

$$z_{2s} = -\frac{1}{c}\left(z_{1s}^3 + az_{1s}^2 + \lambda z_{1s} + u\right) \tag{31a}$$

$$-\frac{1}{c^3}\left(z_{1s}^3 + az_{1s}^2 + \lambda z_{1s} + u\right)^3 - \frac{\mu}{c}\left(z_{1s}^3 + az_{1s}^2 + \lambda z_{1s} + u\right) + cz_{1s} + v = 0 \tag{31b}$$

The stability of these solutions is determined by the eigenvalues ω of the Jacobian matrix of (30), which satisfy the characteristic equation

$$\omega^2 + \left\{3\left(z_{1s}^2 + z_{2s}^2\right) + 2az_{1s} + \lambda + \mu\right\} w + \left(3z_{1s}^2 + 2az_{1s} + \lambda\right)\left(3z_{2s}^2 + \mu\right) - c^2 = 0 \tag{32}$$

We notice that Eq. (31b) is of ninth degree. It can thus have up to nine real solutions continuing those of the uncoupled case, as long as c is sufficiently small. The situation will change as c is gradually increased. Since there cannot be more than nine solutions, one expects that the number of solutions will decrease. Furthermore, degenerate situations where two real solutions were merging for $c = 0$ should now give rise to two structurally stable solutions following the splitting of the degeneracy caused by the "perturbation" terms $c\bar{z}_2$ and $c\bar{z}_1$.

Figure 1 depicts the bifurcation diagram of solution z_{1s} [Eq. (31b)] versus parameter c keeping a, λ, μ, u, and v fixed such that in the limit $c = 0$ one has nine real solutions. As can be seen, increasing c leads successively from four simultaneously stable solutions (among a total of nine solutions) to three simultaneously stable solutions (among a total of seven solutions) and finally to two simultaneously stable solutions (among a total of three solutions). This is in full agreement

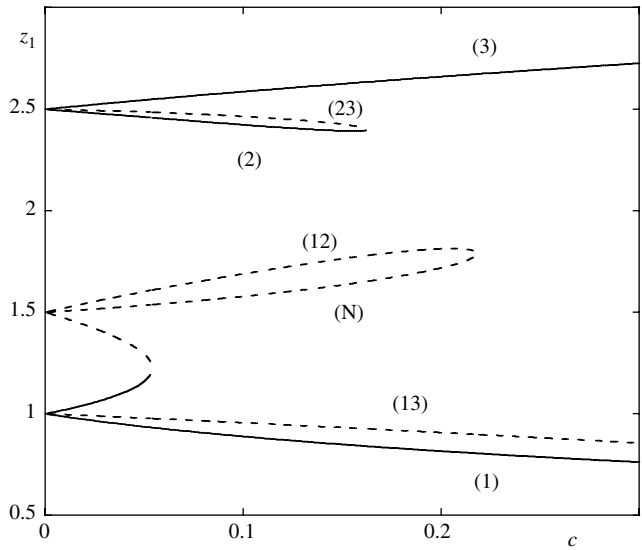


Figure 1. Bifurcation diagram of solution z_{1s} of Eq. (31b) versus parameter c for $a = -5$, $\lambda = 7.75$, $\mu = -7/4$, $u = -3.75$, and $v = 1/2$. Full and dashed lines denote stable and unstable states, respectively. (1) Reference stable state; (3) final stable state; (2) intermediate stable state; (12), (13), (23) intermediate unstable saddle-type states; (N) unstable node.

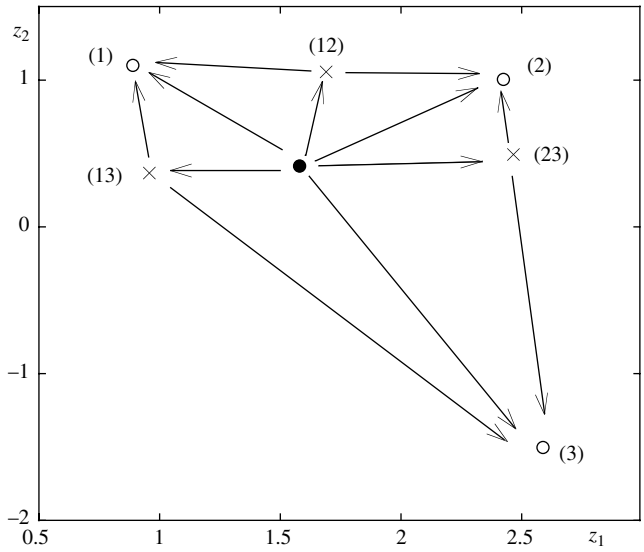


Figure 2. Location of steady states of Eq. (30) and of the unstable manifolds in the two-dimensional phase space (z_1, z_2) for $c = 0.1$. Other parameter values as in Fig. 1.

with the qualitative arguments advanced above. Notice that from the standpoint of phase transition theory four coexisting stable phases is compatible with the Gibbs phase rule [30] as long as one is dealing with a two-component system, as is the case of a protein solution.

The location of the steady states in phase space in the range of three simultaneously stable states is indicated in Fig. 2, where the unstable manifolds of the other fixed points are also sketched. We obtain a configuration corresponding exactly to what was anticipated in the beginning of Section IVB showing clearly the existence of two *a priori* competing pathways, of transitions from state 1 to state 3, $1 \to (13) \to 3$ and $1 \to (12) \to 2 \to (23) \to 3$. This configuration also suggests associating in the particular case of protein crystallization order parameter z_1 to "concentration" and order parameter $-z_2$ to "structure" or "crystallinity."

Naturally, the transitions between the stable states present in the bifurcation diagram of Fig. 1 will be governed by the augmented version of Eq. (30) in which fluctuations are incorporated, see Eq. (18a):

$$\frac{dz_1}{dt} = -z_1^3 - az_1^2 - \lambda z_1 - cz_2 - u + r_1(t)$$
$$\frac{dz_2}{dt} = -z_2^3 - \mu z_2 - cz_1 - v + r_2(t) \tag{33}$$

where the random forces r_1, r_2 satisfy conditions (18b).

V. KINETICS OF TRANSITIONS BETWEEN STATES: MAPPING INTO A DISCRETE MARKOV PROCESS

As stressed in Sections II and III, a case of special interest when studying transitions between states is the weak noise limit. For a system whose evolution is generated by a kinetic potential U, in the notation of Eq. (18b) and in the light of the comments following Eq. (20), this means that ϵ is much smaller than the various potential barriers ΔU_{ij}. The latter are given by the differences of the value $U(ij)$ of U on the unstable (transition) state (ij) across which the system will leave attraction basin of stable state i to enter that of stable state j, and of the value $U(i)$ of U on the reference stable state i.

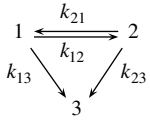
The mean waiting time τ_{ij} for a transition from state i across state (ij) is given by an extension of a classic theory originally elaborated for one-variable systems by Kramers [2, 3] in the multi (here two-) dimensional case. In what follows we will be especially interested in the corresponding transition rates $k_{ij} = \tau_{ij}^{-1}$ that in the above specified limits take the form [11, 31]

$$k_{ij} = \frac{1}{2\pi}\left(\sigma_i^{(1)}\sigma_i^{(2)}\right)^{1/2}\left(\frac{\sigma_{(ij)}^{+}}{|\sigma_{(ij)}^{-}|}\right)^{1/2}\exp\left(-\frac{\Delta U_{ij}}{\epsilon}\right) \tag{34}$$

Here $\sigma_i^{(1)}, \sigma_i^{(2)}$ are the eigenvalues of the Hessian of U (or equivalently the solutions of Eq. (32) on the stable reference state i); and $\sigma_{(ij)}^{\pm}$ stand, respectively, for the unstable and stable eigenvalue of the Hessian [or equivalently the solutions of Eq. (32)] evaluated on the unstable (saddle) transition point.

Under the same assumptions the diffusion type of stochastic process described by Eq. (33) can be mapped into a jump process between the (discrete) stable states, in which the transition probabilities per unit time are given by Eq. (34), see Refs 2, 26, 27. This leads to the following simple kinetic schemes, it being understood that state 3 is by far the most stable state and thus plays the role of an absorbing barrier.

A. Three Simultaneously Stable States (Seven-Steady-State Region)



(35)

The pobabilities p_1 and p_2 to be in the attraction basins of states 1 and 2 satisfy the rate equations

$$\begin{aligned}\frac{dp_1}{dt} &= -(k_{12}+k_{13})p_1 + k_{21}p_2\\ \frac{dp_2}{dt} &= k_{12}p_1 - (k_{21}+k_{23})p_2\end{aligned} \tag{36a}$$

with

$$p_3 = 1 - p_1 - p_2 \tag{36b}$$

The overall transition rates from 1 to 3 will be determined by the eigenvalues $\omega_\pm$ of the transition matrix associated to Eq. (36a),

$$\omega_\pm^2 + (k_{12} + k_{21} + k_{23} + k_{13})\omega_\pm + k_{12}k_{23} + k_{21}k_{13} + k_{23}k_{13} = 0$$

Actually, following a short initial transient the dominant time scale of the transition will be determined by the smallest (in absolute value) eigenvalue ω_-,

$$\omega_- = \frac{1}{2}\{-(k_{12} + k_{21} + k_{23} + k_{13}) + \sqrt{(k_{12} + k_{21} + k_{23} + k_{13})^2 - 4(k_{12}k_{23} + k_{21}k_{13} + k_{23}k_{13})}\} \tag{37}$$

B. Two Simultaneously Stable States (Five-Steady-State Region)

$$1 \underset{k'_{13}}{\overset{k_{13}}{\rightleftharpoons}} 3 \tag{38}$$

This configuration arises past the limit point where states (23) and 2 merge in the bifurcation diagram of Fig. 1. There are two independent pathways leading from 1 to 3, a first past unstable state (13) exactly as in case A and a second past unstable state $(13)'$ that is actually a smooth continuation of unstable state (12) of case A. The probability p_1 to be in the attraction basin of state 1 satisfies the equation

$$\frac{dp_1}{dt} = -(k_{13} + k'_{13})p_1 \tag{39a}$$

with

$$p_3 = 1 - p_1 \tag{39b}$$

The overall transition rate from 1 to 3 is here determined by

$$|\omega| = (k_{13} + k'_{13}) \tag{40}$$

C. Two Simultaneously Stable States (Three-Steady-State Region)

This configuration arises past the limit point where states (12) and N merge in the bifurcation diagram of Fig. 1. It corresponds to the simple kinetic scheme

$$1 \xrightarrow{k_{13}} 3 \tag{41}$$

with

$$\frac{dp_1}{dt} = -k_{13}p_1 \tag{42a}$$

$$p_3 = 1 - p_1 \tag{42b}$$

The transition rate is simply determined by

$$|\omega| = k_{13} \tag{43}$$

Equations (37), (40), and (43) in conjunction with Eqs. (34), (29), and (31) allow one to determine the mean transition times $\tau_{13} = 1/|\omega_-|$, $1/(k_{13} + k'_{13})$, and $1/k_{13}$ as the parameter c runs from left to right across the bifurcation diagram of Fig. 1. The result is represented by the curve in full line in Fig. 3. For reference we also plot (dashed line) the transition times $\tau_{13}^{(0)}$ from 1 to 3 ignoring the presence or not of state 2, evaluated as $\tau_{13}^{(0)} = k_{13}^{-1}$. Finally, the crosses in Fig. 3 stand for the result of a simulation of the full stochastic differential equations (33). As can be seen the analytic and the simulation results agree remarkably well. Deviations do occur but they are limited to situations near transition points where the potential barrier is small, a case for which Eq. (34) needs to be amended as one of the assumptions of the theory ($\epsilon \ll \Delta U$) is no longer fulfilled. More importantly, we see that as one is entering the region of existence of stable state 2 the transition times decrease dramatically as compared to those that would prevail had the direct path $1 \to 3$ via unstable state (13) been the only one available. In particular, for values of c slightly less that 0.1 state 2 turns out to be less stable than state 1 and the barrier ΔU_{23} less than ΔU_{13}. The transition from 1 to 3 occurs then in an accelerated way through a "secondary nucleation" mediated by state 2. Remarkably, state 2 has an accelerating effect even in the vicinity of the transition point where it merges with unstable state (23). This occurs near $x \approx 0.16$ in the bifurcation diagram of Fig. 1 and may be referred to as "transient intermediate state nucleation," see also chapter by J. Lutsko in this volume. These conclusions are in accord with the results of observations [13], simulations [12], and microscopic approaches [14, 15] according to which the presence of an intermediate dense fluid phase enhances the nucleation of protein crystals. It also shows that intermediate state-mediated

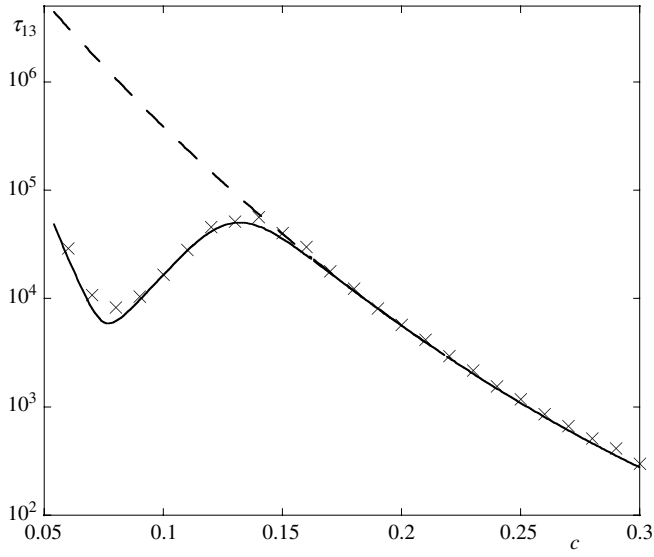


Figure 3. Mean transition times τ_{13} between states 1 and 3 versus parameter c. Full line: analytic result [Eqs. (34) and (37)]; crosses: result of the simulation of the full stochastic differential equations (33); dashed line: direct transition from 1 to 3 ignoring the presence of state 2. Noise strength $2\epsilon = 0.025$. Number of realizations is 1000 and other parameter values as in Fig. 1.

enhancement is a generic phenomenon, extending beyond the specific case of protein crystallization.

VI. IRREVERSIBLE THERMODYNAMICS OF FLUCTUATION-INDUCED TRANSITIONS

According to the setting adopted throughout this chapter, the transition between the initial state 1 and the final state 3 (direct or indirect via state 2) is a transient irreversible process: starting from state 1 with probability equal to one the state is gradually depleted and eventually state 3 invades the entire system, owing to its higher stability ($U(3) \ll U(1)$) in conjunction with the weak noise limit $\epsilon < \Delta U_{13} \ll \Delta U_{31}$. In this section, we analyze this transition from the standpoint of irreversible thermodynamics by evaluating the entropy production [8, 9] associated to the different scenarios considered in the previous sections.

As a reference we start with the direct transition scenario (cf. Section VC)

$$1 \underset{k_{31}}{\overset{k_{13}}{\rightleftharpoons}} 3 \tag{44}$$

where we introduce a small backward transition rate ($k_{31} \ll k_{13}$) in order to avoid the complication arising from the presence of purely irreversible steps. We also adopt a "chemical" interpretation where the variables associated to 1 and 3 are the mass fractions x and s of the material in these states [see also comment prior to Eq. (11)]. The rate equation associated to scheme (44) is then [to be compared with Eq. (42)]

$$\frac{dx}{dt} = -k_{13}x + k_{31}s \tag{45a}$$

with

$$s = 1 - x \tag{45b}$$

The solutions of Eqs. (45) corresponding to the initial conditions $x(0) = 1$, $s(0) = 0$ read

$$\begin{aligned} x(t) &= \frac{k_1'}{k_1 + k_1'} + \frac{k_1}{k_1 + k_1'} \exp[-(k_1 + k_1')t] \\ s(t) &= \frac{k_1}{k_1 + k_1'} - \frac{k_1}{k_1 + k_1'} \exp[-(k_1 + k_1')t] \end{aligned} \tag{46}$$

The instantaneous entropy production (normalized by the gas constant) associated to this "direct" transformation is

$$\sigma_{\mathrm{d}}(t) = (k_{13}x - k_{31}s) \ln \frac{k_{13}x}{k_{31}s}$$

where for simplicity we adopted the assumption of an ideal mixture. Integrating this expression from time $t = 0$ to $t = \infty$ and using Eq. (46) one obtains the total dissipation undergone by the system to complete the transition,

$$\Sigma_d = \int_0^\infty dt \sigma_{\mathrm{d}}(t) = \ln \frac{k_{13} + k_{31}}{k_{31}} = \ln(K_{\mathrm{eq}} + 1) \tag{47}$$

where $K_{\mathrm{eq}} = k_{13}/k_{31}$ is the equilibrium constant of the "reaction" in Eq. (44). In the limit $k_{31} \ll k_{13}$ this expression reduces to $\ln K_{\mathrm{eq}}$, that is, to the standard free energy change associated to the reaction.

We next evaluate the dissipation in the "secondary nucleation" scenario, where the 1 to 3 transition occurs via state 2. The corresponding scheme, rate equations and entropy production are (compare with Section VA)

$$1 \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} 2 \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} 3 \tag{48}$$

$$\begin{aligned}\frac{dx}{dt} &= -k_{12}x + k_{21}y \\ \frac{dy}{dt} &= k_{12}x - (k_{21} + k_{23})y + k_{32}s \\ s &= 1 - x - y\end{aligned} \tag{49}$$

$$\sigma = (k_{12}x - k_{21}y)\ln\frac{k_{12}x}{k_{21}y} + (k_{23}y - k_{32}s)\ln\frac{k_{23}y}{k_{32}s}$$

Integrating again σ from 0 to ∞ we obtain for $x(0) = 1$ and in the limit of small k_{21}, k_{32} the following expression for the total dissipation during the transformation:

$$\Sigma = \Sigma_1 + \Sigma_2 \tag{50a}$$

with

$$\Sigma_1 = \ln\frac{k_{12}}{k_{21}} + \ln\frac{k_{23}}{k_{32}} \tag{50b}$$

and

$$\begin{aligned}\Sigma_2 = -k_{12}\int_0^\infty dt\exp[-k_{12}t]\ln\frac{k_{12}(1-\exp[-(k_{23}-k_{12})t])}{k_{23}-k_{12}} + \frac{k_{12}k_{23}}{k_{12}-k_{23}} \\ \times\int_0^\infty dt(\exp[-k_{12}t] - \exp[-k_{23}t]) \\ \times\ln\frac{k_{23} - k_{12} - k_{23}\exp[-k_{12}t] + k_{12}\exp[-k_{23}t]}{k_{12}(\exp[-k_{12}t] - \exp[-k_{23}t])}\end{aligned} \tag{50c}$$

Here part Σ_1 corresponds to the contribution of the standard part of the chemical potentials in the entropy production and part Σ_2 to the contribution of the composition-dependent part. Clearly,

$$\Sigma_1 = \ln\frac{k_{12}k_{23}}{k_{21}k_{32}} = \ln K_{\text{eq}}$$

where K_{eq} was already introduced in Eq. (47). In the limit of small k_{32}, Σ_1 becomes thus identical to Σ_d. On the other hand, as it turns out $\Sigma_2 = 0$ for all values of the ratio k_{23}/k_{12}. We therefore conclude that

$$\Sigma = \Sigma_d \tag{51}$$

that is, at least in the limits considered the total dissipation associated to transition from 1 to 3 is independent of whether the transition is direct or occurs via the intermediate state 2. This is at variance with some results reported in the literature in a somewhat different context [32], where the presence of intermediate steps

seems to lower the entropy production. What it happening here is, rather, that the contribution to dissipation due to the distance from equilibrium of the ratio x/y counteracts exactly that due to the distance from equilibrium of the ratio y/s. We emphasize that nonequilibrium is manifested here as a transient phenomenon, as eventually the whole system is invaded with probability one by the single final state 3, whatever the pathway for reaching this state might have been.

VII. CONCLUSIONS

In this chapter we outlined a general approach for analyzing the transitions between simultaneously stable steady states in the presence of thermodynamic or externally driven fluctuations. Emphasis was placed on transitions between two given states taking place under conditions allowing for the presence of a third one, which could be metastable with respect to the other states or even appear as a transient in the vicinity of criticalities of certain kinds. Conditions were defined under which the presence of such "intermediate" states can enhance the rate of transitions between the two "reference" states. We have shown that these mechanisms are likely to be generic and underlie a wide range of phenomena including the nucleation of crystalline materials from a liquid mixture in the presence of a second liquid phase as observed, in particular, in the crystallization of globular proteins or colloids.

Throughout the analysis we considered a closed system evolving eventually toward the state of thermodynamic equilibrium and satisfying the condition of detailed balance, entailing a vanishing probability flux in each phase space point and the proportionality between the Onsager matrix of kinetic coefficients and the diffusion matrix of the fluctuations. Under the additional condition that these matrices are state independent we showed that the evolution of the state variables could be mapped into a form deriving from a kinetic potential. We have constructed generic potentials compatible with various nucleation scenarios, some of which are identical to experimentally observed scenarios in the context of crystallization of globular proteins or colloids, see Ref. 13 and the Chapters by Vekilov and by Dinsmore et al. in this volume. These potentials provide also valuable information on the nature of the free energy surface of the system concerned which, despite its complex dependence on microscopic-level properties, is ultimately topologically equivalent to them.

More than one century ago, Wilhelm Ostwald proposed a *step rule* to explain certain observations on the formation of crystalline materials from the melt [33]. According to this rule, typically it is not the most stable form of the material but the least stable one that crystallizes first. The status of this rule in the light of present day knowledge on nucleation has been discussed recently in the literature [34, 35]. Our analysis shows that Ostwald's step rule is a consequence of a general underlying mechanism: the transition from an initial state (a liquid phase) to a final

one (a stable crystalline phase) may occur preferably via a pathway involving an intermediate state (a second, metastable liquid phase), possessing the necessary properties (e.g., high concentration of a key substance) for the appearance and growth of nuclei possessing the structure of the final phase. The analysis also shows that this is only one out of several scenarios and provides information on its realizability depending on the values of the parameters present in the problem.

Transitions between states are a generic phenomenon observed in a multitude of systems of interest in physical, life, engineering, and environmental sciences [36]. It constitutes one of the principal signatures of the dynamics of complex systems as it endows them with the ability to evolve, to choose, and to adapt and extends in this respect beyond the examples of protein or colloid crystallization and similar phase transition phenomena. Viewed in all its generality it constitutes a largely open problem likely to witness important developments in the future.

A first direction would be to account for the presence of several intermediate states as it may be the case, for instance, in transitions between different conformations of proteins and other biomolecules [37].

From the standpoint of stochastic processes and thermodynamics an interesting extension would be to assess whether some of our conclusions subsist when the condition of detailed balance breaks down. There are two broadly different mechanisms at the origin of such a breakdown.

(i) The evolution vector $\mathbf{v}$ in Eq. (1) contains an additional contribution beyond the one involving the thermodynamic or the kinetic potential. Typically, this will reflect the presence of constraints in, for instance, the form of concentration or temperature gradients, that do not allow the system to reach the state of equilibrium as provided by the extremum of the potential, even in the long time limit.

(ii) The diffusion matrix D_{ij} of the fluctuations accounts for the fact that the medium in which the system of interest is embedded is itself maintained out of equilibrium—for instance, by an external time-periodic driving.

In both cases, while Eqs. (6) and (7) will remain valid Eqs. (18) will break down and much of the analysis of Sections III–V will need to be extended. In particular, the exit from the attraction basin of a stable state need no longer follow a path across an intermediate unstable state of the saddle point type.

The influence of intermediate states on the kinetics of the transitions between two reference states can also be viewed as a new mechanism of control and optimization of the rate of production of the "material" associated to the final state. In this perspective, it would be of interest to study its interference with other known control mechanisms. Most prominent among them is stochastic resonance [38, 39], whereby the passage over a potential barrier is facilitated by the presence of a weak external periodic forcing. In the context of phase transitions and in particular

of protein crystallization a coupling of nucleation kinetics with an external field is expected to exist and play a nontrivial role in view of the presence of charged groups in the individual molecules [26]. Ordinarily, studies in stochastic resonance focus on the case of two simultaneously stable states separated by an unstable state of the saddle point type. Recently, an extension to the case of several stable states in the presence of a spatially periodic potential has been reported [40]. It would undoubtedly be interesting to adapt this work to the type of potentials considered in this chapter and assess the additional effect of the external periodic forcing in different parts of the bifurcation diagram of Fig. 1.

Finally, future developments in this area should aim at the elucidation of the microscopic basis of the mesoscopic-level approach developed in this chapter. This should clarify the limits of validity of some of the assumptions adopted, such as the use of spatially lumped variables and the independence of matrices L and D on the state variables $\{x_i\}$ in Eqs. (14) and onward, and provide the basis for appropriate refinements and generalizations.

Acknowledgments

We acknowledge a fruitful discussion with Professor R. McKay. This work is supported, in part, by the European Space Agency and the Belgian Federal Science Policy Office under contract numbers C90238 and C90241.

References

1. G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems*, Wiley, New York, 1977; H. Haken, *Synergetics: A Introduction*, Springer, Berlin, 1977.
2. C. Gardiner, *Handbook of Stochastic Methods*, Springer, Berlin, 1983.
3. N. van Kampen, *Stochastic Processes in Physics and Chemistry*, North Holland, Amsterdam, 1981.
4. K. Tomita, T. Ohta, and H. Tomita, *Prog. Theor. Phys.*, **52**, 1744 (1974).
5. R. Kubo, K. Matsuo, and K. Kitahara, *J. Stat. Phys.*, **9**, 51 (1973).
6. H. Gang, *Phys. Rev. A*, **36**, 5782 (1987).
7. L. Bertini, A. D. Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, *Phys. Rev. Lett.*, **87**, 040601 (2001).
8. P. Glansdorff and I. Prigogine, *Thermodynamic Theory of Structure, Stability and Fluctuations*, Wiley, London, 1971.
9. S. De Groot and P. Mazur, *Nonequilibrium Thermodynamics*, North Holland, Amsterdam, 1962.
10. G. Nicolis and C. Nicolis, *Phys. Rev. E*, **77**, 051101 (2008).
11. B. Matkowsky and Z. Schuss, *SIAM J. Appl. Math.*, **33**, 365 (1977).
12. P. ten Wolde and D. Frenkel, *Science*, **277**, 1975 (1997).
13. P. Vekilov, *Cryst. Growth Des.*, **4**, 671 (2004).
14. V. Talanquer and D. Oxtoby, *J. Chem. Phys.*, **109**, 223 (1998).
15. J. Lutsko and G. Nicolis, *Phys. Rev. Lett.*, **96**, 046102 (2006).
16. L. Landau and E. M. Lifshitz, *Statistical Physics*, Pergamon, Oxford, 1959.

17. N. Boccara, *Symétries Brisées*, Hermann, Paris, 1976.
18. J. Guckenheimer and Ph. Holmes, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, Springer, Berlin, 1983.
19. G. Nicolis, *Introduction to Nonlinear Science*, Cambridge University Press, Cambridge, 1995.
20. R. Thom, *Stabilité Structurelle et Morphogénèse*, Benjamin, New York, 1972.
21. P. Saunders, *An Introduction to Catastrophe Theory*, Cambridge University Press, Cambridge, 1980.
22. I. Stewart, *Physica*, **2D**, 245 (1981).
23. K. Okada, *Prog. Theor. Phys.*, **78**, 1236 (1987).
24. J. Gaite, *Phys. Rev. A*, **41**, 5320 (1990).
25. J. Gaite, J. Margalef-Roing, and S. Miret-Artés, *Phys. Rev. B*, **59**, 8593 (1999).
26. G. Nicolis and C. Nicolis, *Physica A*, **323**, 139 (2003).
27. G. Nicolis and C. Nicolis, *Physica A*, **351**, 22 (2005).
28. V. I. Arnold, *Russ. Math. Surveys*, **30**, 1 (1975).
29. J. Callahan, *Proc. London Math. Soc.*, (3) **45**, 227 (1982).
30. J. W. Mullin, *Crystallization*, Butterworth-Heinemann, London, 1997.
31. P. Hänggi, P. Talkner, and M. Borkovec, *Rev. Mod. Phys.*, **62**, 251 (1990).
32. R. van Santen, *J. Phys. Chem.*, **88**, 5768 (1984).
33. W. Ostwald, *Z. Phys. Chem.*, **22**, 289 (1897).
34. P. ten Wolde and D. Frenkel, *Phys. Chem. Chem. Phys.*, **1**, 2191 (1999).
35. Z. Tavassoli and R. Sear, *J. Chem. Phys.*, **116**, 5066 (2002).
36. G. Nicolis and C. Nicolis, *Foundations of Complex Systems*, World Scientific, Singapore, 2007.
37. C. Wagner and T. Kiefhaber, *Proc. Nat. Acad. Sci. USA*, **96**. 6716 (1999).
38. C. Nicolis, *Tellus*, **34**, 1 (1982).
39. L. Gammaitoni, P. Hänggi, P. Jung and F. Marchesoni, *Rev. Mod. Phys.*, **70**, 223 (1998).
40. C. Nicolis, *Phys. Rev. E*, **82**, 011139 (2010).

DYNAMICAL RARE EVENT SIMULATION TECHNIQUES FOR EQUILIBRIUM AND NONEQUILIBRIUM SYSTEMS

TITUS S. VAN ERP

Centrum voor Oppervlaktechemie en Katalyse, K.U. Leuven, Kasteelpark Arenberg 23, B-3001 Leuven, Belgium

CONTENTS

I. INTRODUCTION

Molecular dynamics (MD) is the ultimate method to gain detailed atomistic information of dynamical processes that are difficult to access experimentally. However, an important bottleneck of atomistic simulations is the limited system- and timescales. Depending on the complexity of the forcefields (*ab initio* MD being extremely more expensive than classical MD) systems typically consist of 100 to 100,000 molecules that can be simulated for a period of nanoseconds

Advances in Chemical Physics, Volume 151: Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanoscale Materials, First Edition. Edited by Gregoire Nicolis and Dominique Maes.

till microseconds. Therefore, many activated processes cannot be studied using brute-force MD because the probability to observe a reactive event within reasonable CPU time is basically zero. Typical examples are protein folding, conformational changes of molecules, cluster isomerizations, chemical reactions, diffusion in solids, ion permeation through membranes, enzymatic reactions, docking, nucleation, DNA denaturation, and other types of phase transitions. If these processes are treated with straightforward MD, the simulations will endlessly remain in the reactant states. Still, if an event would happen, it can go very fast. The time it actually takes to cross the barrier is usually much shorter than this computational accessible timescale. Therefore, rare event algorithms aim to avoid the superfluous exploration of the reactant state and to enhance the occurrence of reactive events. The methods that I will discuss are the reactive flux (RF) method [1, 2] and the more recent algorithms that originate from the transition path sampling (TPS) [3–7] methodology. These comprise the transition interface sampling (TIS) [8] and the replica exchange TIS (RETIS) [9, 10], which are successive improvements on the way reaction rates were determined in the original TPS algorithm. Partial path TIS (PPTIS) [11] is an approximative approach in order to reduce the simulated path length for the case of diffusive barrier crossings. PPTIS is similar to Milestoning [12], that was developed simultaneously and independently from PPTIS. For nonequilibrium systems, the forward flux sampling (FFS) was designed [13]. This method is based on the TIS formalism, but does not require prior knowledge on the phasepoint density. All these methods have in common that they aim to simulate true molecular dynamics trajectories at a much faster rate than naive brute force molecular dynamics. I will discuss the advantages and disadvantages of the different methodologies and introduce a few new relations and derive some known relations using a nonstandard approach. The descriptions of these methods given here are far from complete and, therefore, to obtain a more complete picture of the path sampling techniques I would like to recommend some very recent complementary reviews on these methodologies [14–17]. In the end, I compare all the methods by applying them on a simple, though tricky, test system. The outcome illustrates some important pitfalls for the nonequilibrium methods that have no easy solution and show that caution is necessary when interpreting their results.

II. REACTIVE FLUX METHOD

Low dimensional systems, such as chemical reactions in the gas phase, are usually well described by transition state theory (TST). TST assumes that the transitions from reactant to product state always follow a path on the potential energy surface such that it passes the barrier nearby the transition state (TS). The TS refers to the point on the energy barrier having to the lowest possible potential energy difference, with respect to the reactant state, that any trajectory must overcome in order to reach

the product state. In this description, TS corresponds to a unstable stationary point on the potential energy surface having one imaginary frequency (saddle point). In condensed systems, the saddle point of the full potential energy surface is usually less meaningful. For instance, if we would consider the dissociation of NaCl in water, the TS would correspond to a state where the interion distance is fixed to a critical value while all surrounding water molecules are full frozen into an icy state. It is needless to say that this does not correspond to our daily experience when dissolving some pinch of salt in a glass of water. TST can be generalized for higher dimensional systems using the free energy instead of the potential energy. The TS is then no longer a single point, but a multidimensional surface. In this case, the TST equation is determined by the free energy difference between the TS dividing surface and the reactant state. An important limitation of TST is that the free energy barrier depends on the selected degrees of freedom that are used to describe the free energy surface. In addition, TST implicitly assumes that any trajectory will cross the free energy barrier only once when going from reactant to product state. Kramers' theory [18] provides an elegant and insightful approach to correct for correlated recrossings if these originate from the diffusive character of the dynamics. However, there are several other sources for recrossings. For instance, if the selected degrees of freedom are not well chosen, the barriers in the free energy landscape do not always correspond the barriers in the underlying potential energy surface which ultimately determine the dynamics. The RF method is able to correct for recrossings regardless their origin and is very powerful when TST is close but not sufficiently accurate.

The theory of the method originated from the early 1930s, far before the first applications of computers for molecular dynamics simulations [19]. After Wigner and Eyring introduce the concept of the TS and the TST approximation [20, 21], Keck [22] demonstrated how to calculate the dynamical correction, the transmission coefficient. This work has later been extended by Bennett [23], Chandler [24], and others [25, 26], resulting in a two-step approach. First the free energy as function of a single reaction coordinate (RC) is determined. This can be done by, for example, umbrella sampling (US) [27] or thermodynamic integration (TI) [28]. Then, the maximum of this free energy profile defines the approximate TS dividing surface and the transmission coefficient can be calculated by releasing dynamical trajectories from the top.

Traditionally, the equation for the dynamically corrected rate constant is derived by applying a small perturbation to the equilibrium state and invoking the fluctuation–dissipation theorem and Onsager's relation [1, 2, 25]. However, as I will show here, there is an alternative derivation that naturally evolves to a formula for transmission coefficient that is probably more efficient that the standard one [23, 24].

There are several definitions for the rate constant k_{AB} between two states A and B, such as the transition probability per unit time, the inverse mean residence

time in state A, or the inverse mean first passage time toward state B [29]. However, all these different definitions become equivalent for truly exponential relaxation, which is the case whenever the stable states A and B are separated by large free energy barriers. If this is not the case, the rate constant becomes ill-defined. To start the derivation I will use the first definition, which can be expressed as follows:

$$k_{AB} = \lim_{dt\to 0} \frac{1}{dt} \frac{\text{number of states } A \text{ that transform into state } B \text{ within } dt}{\text{number of states } A} \tag{1}$$

Let us denote $x = (r, v)$ the phasepoint which includes the positions r and velocities v of all particles in the system. We define the reaction coordinate $\lambda(x)$ that can be any function of x, though in practice it will generally only depend on r. The RC function should describe the progress of the reaction, but there is a lot of flexibility in designing this RC function.

We will assume that the collection of phasepoints $\{x|\lambda(x) = 0\}$ defines the transition state dividing surface that separates region A and B. For convenience, we will also assume that the RC will increase when going from A to B. Considering the phenomenological equation (1), we can directly write down the reaction rate as

$$\begin{aligned} k_{AB} &= \lim_{dt\to 0} \frac{1}{dt} \frac{\int dx_0\, \theta\left(-\lambda\left(x_0\right)\right) \theta\left(\lambda\left(x_{dt}\right)\right) \rho(x_0)}{\int dx_0 \theta\left(-\lambda\left(x_0\right)\right) \rho(x_0)} \\ &= \lim_{dt\to 0} \frac{1}{dt} \frac{\left\langle \theta\left(-\lambda\left(x_0\right)\right) \theta\left(\lambda\left(x_{dt}\right)\right)\right\rangle}{\left\langle \theta\left(-\lambda\left(x_0\right)\right)\right\rangle} \end{aligned} \tag{2}$$

where x_0 and x_{dt} are phasepoints at times $t = 0$ and $t = dt$. $\rho(x)$ denotes the phasepoint density. For equilibrium statistics this is simply given by Boltzmann $\rho(x) = \exp(-\beta E(x))$, where E the energy and $\beta = 1/k_{\mathrm{B}}T$, T the temperature, and k_{B} the Boltzmann constant. θ is the Heaviside-step function with $\theta(y) = 0$ if $y < 0$ and $\theta(y) = 1$ otherwise. The brackets $\langle \ldots \rangle \equiv \int dx \ \ldots \rho(x) / \int dx\, \rho(x)$ denote the ensemble average over the initial condition x_0. Equation (2) is basically the TST expression of the rate, but written in a somewhat unusual form.

To transform this equation into the standard form, we can use $\lambda\left(x_{dt}\right) = \lambda\left(x_0\right) + dt\dot{\lambda}(x_0) + \mathcal{O}(dt^2)$, where the dot denotes the time derivative. If we neglect the second order terms, we can write for an arbitrary function $a(x)$:

$$\int dx_0\, \theta\left(-\lambda\left(x_0\right)\right) \theta\left(\lambda\left(x_{dt}\right)\right) a(x_0) = \int dx_0\, \theta\left(-\lambda\left(x_0\right)\right) \theta\left(\lambda\left(x_0\right) + dt\dot{\lambda}\left(x_0\right)\right) a(x_0) \tag{3}$$

Clearly, $\theta(-\lambda)\theta(\lambda + dt\dot{\lambda})$ is only nonzero if $\dot{\lambda} > 0$ and $dt\dot{\lambda} < \lambda < 0$. Instead of integrating over x_0, we will apply a coordinate transform such that we can integrate Eq. 3 over $\lambda, \dot{\lambda}$ and a remaining set coordinates x_0'. Assume that $J(x_0', \lambda, \dot{\lambda})$ is the

corresponding Jacobian of this transformation. We can then integrate out the $(\lambda, \dot{\lambda})$ coordinates

$$\begin{aligned}
&\int dx_0\, \theta(-\lambda(x_0))\, \theta\left(\lambda(x_0) + dt\dot{\lambda}(x_0)\right) a(x_0) \\
&= \int dx_0' \int_0^\infty d\dot{\lambda} \int_{-dt\dot{\lambda}}^0 d\lambda\, a(x_0', \lambda, \dot{\lambda}) J(x_0', \lambda, \dot{\lambda}) \\
&= \int dx_0' \int_0^\infty d\dot{\lambda} \int_{-dt\dot{\lambda}}^0 d\lambda \left\{ a(x_0', 0, \dot{\lambda}) J(x_0', 0, \dot{\lambda}) \right. \\
&\quad \left. + \lambda \frac{\partial (a(x_0', \lambda, \dot{\lambda}) J(x_0', \lambda, \dot{\lambda}))}{\partial \lambda}\Big|_{\lambda=0} + \cdots \right\} \\
&= \int dx_0' \int_0^\infty d\dot{\lambda} \left\{ (dt\dot{\lambda}) a(x_0', 0, \dot{\lambda}) J(x_0', 0, \dot{\lambda}) \right. \\
&\quad \left. - \frac{1}{2}(dt\dot{\lambda})^2 \frac{\partial (a(x_0', \lambda, \dot{\lambda}) J(x_0', \lambda, \dot{\lambda}))}{\partial \lambda}\Big|_{\lambda=0} + \cdots \right\} \\
&= dt \times \int dx_0' \int_0^\infty d\dot{\lambda}\, \dot{\lambda} a(x_0', 0, \dot{\lambda}) J(x_0', 0, \dot{\lambda}) + \mathcal{O}(dt^2)
\end{aligned} \tag{4}$$

where we applied a Taylor expansion in terms of λ in the second line. Clearly, as

$$\begin{aligned}
&\int dx_0\, \dot{\lambda}(x_0)\, \delta(\lambda(x_0))\, \theta\left(\dot{\lambda}(x_0)\right) a(x_0) \\
&= \int dx_0' d\dot{\lambda} d\lambda\, \dot{\lambda} \delta(\lambda) \theta(\dot{\lambda}) a(x_0', \lambda, \dot{\lambda}) J(x_0', \lambda, \dot{\lambda}) \\
&= \int dx_0' \int_0^\infty d\dot{\lambda}\, \dot{\lambda} a(x_0', 0, \dot{\lambda}) J(x_0', 0, \dot{\lambda})
\end{aligned} \tag{5}$$

we have proven that

$$\lim_{dt \to 0} \frac{1}{dt} \theta(-\lambda(x_0))\, \theta(\lambda(x_{dt})) = \dot{\lambda}(x_0)\, \delta(\lambda(x_0))\, \theta\left(\dot{\lambda}(x_0)\right) \tag{6}$$

Using this expressing into Eq. (2), we obtain the standard form of the *TST formula*:

$$k_{AB} = \frac{\langle \dot{\lambda}(x_0)\, \delta(\lambda(x_0))\, \theta(\dot{\lambda}(x_0)) \rangle}{\langle \theta(-\lambda(x_0)) \rangle} \tag{7}$$

It is often very convenient to switch back to the other formalism, Eq. (2), as some relations follow more naturally from this expression, especially in path sampling simulations where dt can simply be taken as the MD timestep. The TST approach

rewrites Eq. (7) into two factors

$$k_{AB} = \frac{\langle \dot{\lambda}(x_0)\,\delta(\lambda(x_0))\,\theta(\dot{\lambda}(x_0))\rangle}{\langle \delta(\lambda(x_0))\rangle} \times \frac{\langle \delta(\lambda(x_0))\rangle}{\langle \theta(-\lambda(x_0))\rangle} \equiv R^{\mathrm{TST}} \times \frac{e^{-\beta F(0)}}{\int_{-\infty}^{0} d\lambda\, e^{-\beta F(\lambda)}} \tag{8}$$

where the free energy F is defined as $F(\lambda) \equiv -\ln\langle\delta(\lambda - \lambda(x))\rangle/\beta$. Numerous techniques exist to calculate the free energy profile along the barrier region [27, 28, 30–35]. The kinetic term R^{TST} usually follows from a simple numerical or analytical integration. For instance, if the RC is a simple Cartesian coordinate of a target particle then $R^{\mathrm{TST}} = 1/\sqrt{2\pi\beta m}$ where m is the mass of the particle.

The TST expression neglects correlated fast recrossings and, therefore, overestimates the reaction rate. Recrossings can occur due to a diffusive motion on top of the barrier or by kinetic correlations when the kinetic energy of the RC is not dissipated. Another important source of recrossings is when the one-dimensional RC gives an incomplete description of the reaction kinetics [2]. To correct for recrossing we can apply the effective positive flux formalism that neglects the crossings that are not "effective." At each side of the barrier we define regions that are the stable regions A and B. These might be smaller than the regions that we associate to the product and reactant state. Entering A or B implies that the system is committed to that side, that is, it might leave region A or B shortly thereafter, but the chance to rapidly recross the barrier is of the same order as an independent new event. An effective positive crossing is then defined as the first crossing on the trajectory that makes the transition from A to B (see Fig. 1). This leads to the

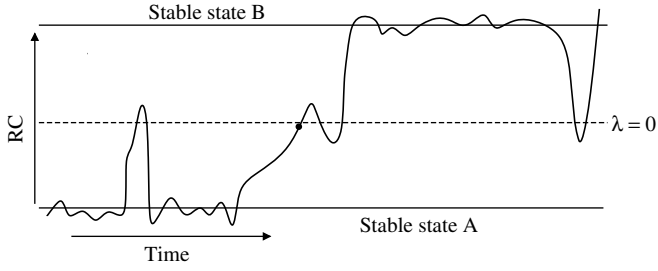


Figure 1. Definition of an effective positive crossing on a very long MD trajectory. The EPF algorithm will ignore all crossing with the TS dividing surface except one (black dot). These are the first crossing points with the TS dividing surface for the parts of the MD trajectory that start at A and end at B (without revisiting A again).

effective positive flux expression for the reaction rate:

$$k_{AB} = \lim_{dt\to 0} \frac{1}{dt} \frac{\langle \theta(-\lambda(x_0))\,\theta(\lambda(x_{dt}))\,h^b_{A0}(x_0)h^f_{BA}(x_0)\rangle}{\langle \theta(-\lambda(x_0))\rangle} = \frac{\langle \dot\lambda(x_0)\,\delta(\lambda(x_0))\,\theta\left(\dot\lambda(x_0)\right)\,h^b_{A0}(x_0)h^f_{BA}(x_0)\rangle}{\langle \theta(-\lambda(x_0))\rangle} \tag{9}$$

where $h^{b/f}_{uv}(x_0)$ detects whether a backward/forward time trajectory crosses or enters a certain interface/region u before interface/region v. If this is true, the function is one. It is zero otherwise. In the second line we applied again equality 6. Naturally, the ensemble average $\langle \ldots \rangle$ should now not only integrate over the phasepoint x_0, but also sum over all possible trajectories backward and forward in time starting from x_0. The ratio between the exact expression [Eq. (9)] and the transition state expression [Eq. (7)] is the transmission coefficient: $k^{\text{exact}} = \kappa k^{\text{TST}}$ so that

$$\kappa = \frac{\langle \dot\lambda(x_0)\delta(\lambda(x_0))\theta(\dot\lambda(x_0))h^b_{A0}(x_0)h^f_{BA}(x_0)\rangle}{\langle \dot\lambda(x_0)\delta(\lambda(x_0))\theta(\dot\lambda(x_0))\rangle} = \frac{\langle \dot\lambda(x_0)\theta(\dot\lambda(x_0))h^b_{A0}(x_0)h^f_{BA}(x_0)\rangle_{\lambda=0}}{\langle \dot\lambda(x_0)\theta(\dot\lambda(x_0))\rangle_{\lambda=0}} \tag{10}$$

Here the subscript $\lambda = 0$ denotes an ensemble average on the TST dividing surface. Strictly speaking, the above expression is correct for any surface that separates the two stable states. However, the efficiency to calculate the above expression is significant better if κ is maximized. Therefore, $\lambda = 0$ should be defined on the top of the free energy barrier. If we assume that $\lambda(x) = \lambda(r)$ depends on configuration space only, the calculation requires to generate a representative set of configuration points on the TST surface. Then, we attribute to these points r a randomized set of velocities taken from a Maxwellian distribution and integrate the equations of motion backward and forward in time. However, as $\theta\left(\dot\lambda\right) h^b_{A0} = 0$ if $\dot\lambda < 0$ or when the backward trajectory recrosses the TST dividing surface before entering A, only a very few trajectories need to be fully integrated in both time directions until reaching stable states. It is surprising that the effective positive flux counting strategy is not so common. To our knowledge only two slightly different expressions of a transmission coefficient based on the effective positive flux have been proposed in Refs 36, 37. All other expressions in the literature do not avoid the counting of recrossings. In these algorithms, the final rate constant follows through cancellation of many negative and positive terms. For instance, the most popular formulation of the rate constant and transmission coefficient is the *Bennett–Chandler (BC)* expression that appears in many textbooks on molecular

simulation [1, 2].

$$\tilde{k}_{AB}(t) = \frac{\langle \dot{\lambda}(x_0)\,\delta(\lambda(x_0))\,\theta(\lambda(x_t))\rangle}{\langle \theta(-\lambda(x_0))\rangle} \Rightarrow$$
$$\tilde{\kappa}(t) = \frac{\langle \dot{\lambda}(x_0)\,\theta(\lambda(x_t))\rangle_{\lambda=0}}{\langle \dot{\lambda}(x_0)\,\theta(\dot{\lambda}(x_0))\rangle_{\lambda=0}} \tag{11}$$

Here, the reaction rate and transmission coefficient are expressed as time-dependent functions. However, the actual rate constant and transmission coefficient, which should not depend on time, follow from a plateau value of these time-dependent functions: $k_{AB} = \tilde{k}_{AB}(t')$, $\kappa = \tilde{\kappa}(t')$ with $\tau_{\mathrm{mol}} < t' < \tau_{\mathrm{rxn}}$. In other words, $\tilde{k}(t)$ and $\kappa(t)$ will generally show oscillatory behavior at small t. However, after some molecular timescale τ_{mol}, the system will basically enter either region A or B (see Fig. 1) after which we would not expect any recrossing until reaching the actual relaxation time $\tau_{\mathrm{rxn}} \gg \tau_{\mathrm{mol}}$. The equivalence between Eq. (11) and Eqs. (9) and (10), can be shown by invoking $\theta(\lambda(x_{t'})) = h^f_{BA}(x_0)$, Eq. (6), and its mirror equivalent $\lim_{dt\to 0} \theta(\lambda(x_0))\theta(-\lambda(x_{dt}))/dt = -\dot{\lambda}(x_0)\delta(\lambda(x_0))\theta(-\dot{\lambda}(x_0))$:

$$\begin{aligned} k = \tilde{k}(t') &= \frac{\langle \dot{\lambda}(x_0)\delta(\lambda(x_0))\theta(\lambda(x_{t'}))\rangle}{\langle \theta(-\lambda(x_0))\rangle} = \frac{\langle \dot{\lambda}(x_0)\delta(\lambda(x_0))h^f_{BA}(x_0)\rangle}{\langle \theta(-\lambda(x_0))\rangle} \\ &= \frac{\langle \dot{\lambda}(x_0)\delta(\lambda(x_0))\theta(\dot{\lambda}(x_0))h^f_{BA}(x_0) + \dot{\lambda}(x_0)\delta(\lambda(x_0))\theta(-\dot{\lambda}(x_0))h^f_{BA}(x_0)\rangle}{\langle \theta(-\lambda(x_0))\rangle} \\ &= \lim_{dt\to 0} \frac{1}{dt}\left(\frac{\langle \theta(-\lambda(x_0))\theta(\lambda(x_{dt}))h^f_{BA}(x_0) - \theta(\lambda(x_0))\theta(-\lambda(x_{dt}))h^f_{BA}(x_0)\rangle}{\langle \theta(-\lambda(x_0))\rangle} \right) \end{aligned} \tag{12}$$

We have now transferred the BC expression in an unitary ensemble average; each phasepoint x_0 either returns 1, 0, or -1. Consider a very long MD trajectory with a timestep of dt (like the one in Fig. 1). It is clear that any detailed-balance simulation method should sample each phasepoint x_0 on this trajectory equally often. As such, an unreactive $B \to B$ trajectory will always have an equal number of phasepoints returning $+1$ as -1. The $B \to B$ trajectories are therefore effectively not counted due to this cancelation. The phasepoints on the $A \to A$ trajectory are always zero due to the h^f_{BA} characteristic function. Finally, any trajectory $A \to B$ always has one x_0 more that is $+1$ than -1. A more formal mathematical proof of the equivalence between Eq. (11) and Eq. (9) can be found in Ref. 38.

Whenever, there are a significant number of recrosssings, the BC formalism has obvious disadvantages. In general, we note that any averaging method counting only zero and positive values will show a faster convergence than one that is based on cancelation of positive and negative terms. Moreover, in the effective flux formalism many trajectories will be assigned as unreactive after just a few MD

steps, thus reducing the number of required force evaluations. Another important advantage of the EPF formalism is that is generates a set of trajectories that are unambiguous interpretable as reactive or unreactive, while the BC scheme generates only forward trajectories of which some actually belong to unreactive $B \rightarrow B$ trajectories. Instead of integrating the equations of motion until reaching stable states, one can also use a time-dependent expression for the EPF [39] similar to Eq. (11).

There are several other formulations of the transmission coefficient [40], but most of them rely on a cancelation between positive and negative flux terms. A comparative study of ion channel diffusion [41] showed that the algorithm based on effective positive flux expression was superior to the other transmission rate expressions. Moreover, it was as efficient as an optimized version of the more complicated method of Ruiz-Montero et al. [42]. The implementation of the EPF scheme is as simple as algorithms that are based on the BC transmission coefficient. Therefore, the EPF implementation of the RF method should in principal be preferred above the standard implementations that require cancelation.

III. TRANSITION PATH SAMPLING

In the previous section, I showed how the standard transmission coefficient calculations can be improved using the effective positive flux expression. However, this approach cannot fully eliminate the main bottleneck of the RF methods. If $\kappa \ll 1$, the number of trajectories that are required for sufficient statistics can be tremendous. In specific, if one is unable to find a proper RC, the overwhelming majority of trajectories that are released from the top of the barrier will be either $A \rightarrow A$ or $B \rightarrow B$ trajectories [2]. In practice, it has been discovered that finding a good RC can be extremely difficult in high dimensional complex systems. Notable examples are chemical reactions in solution, where the reaction mechanism often depends on highly nontrivial solvent rearrangements [43]. Also, computer simulations of nucleation processes use very complicated order parameters to distinguish between particles belonging to the liquid and solid phase. This makes it unfeasible to construct a single RC that accurately describes the exact place of crossover transitions. As a result, hysteresis effects and low transmission coefficients are almost unavoidable.

This has been the main motivation of Chandler and collaborators [3–7] to devised a method that generates reactive trajectories without the need of a RC. This method, called transition path sampling TPS, gathers a collection of trajectories connecting the reactant to the product stable region by employing a Monte Carlo (MC) procedure called *shooting*.

Suppose $\mathbf{x}$ is a path $\{x_0, x_{dt}, x_{2dt}, \ldots, x_{ndt}\}$ of n timeslices. The statistical weight given to this path equals

$$P[\mathbf{x}] = \rho(x_0)p(x_0 \rightarrow x_{dt})p(x_{dt} \rightarrow x_{2dt}) \cdots p(x_{(n-1)dt} \rightarrow x_{ndt})\,\hat{h}(\mathbf{x}) \quad (13)$$

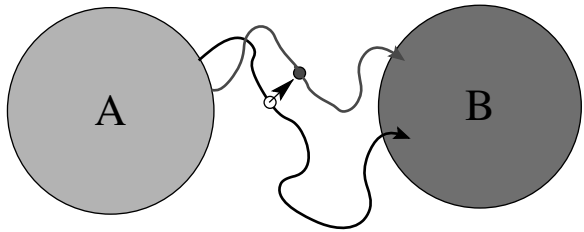


Figure 2. Illustration of the TPS shooting move using flexible path length. From an existing path a random timeslice is selected. Positions and/or velocities of this point are slightly modified giving a new phasepoint. From this point, the equations of motion are integrated forward and backward in time until the trajectory hits A or B.

where $\rho(x_0)$ is the usual phasepoint density and $p(x_{jdt} \to x_{(j+1)dt})$ is probability density that the MD integrator generates $x_{(j+1)dt}$ starting from x_{jdt}. The characteristic function $\hat{h}(\mathbf{x})$ equals 1 (otherwise 0) if a specific condition is fulfilled. For instance, one could imply that the trajectory $\mathbf{x}$ needs to start in state A and end in state B.

By means of the shooting algorithm, TPS performs a random walk in path space to generate one trajectory after the other (see Fig. 2). The first step of this approach consists of a random selection of one of the timeslices of the old path, called the shooting point. This timeslice is modified by making random modifications in the velocities and/or positions. Then, there is usually an acceptance or rejection step based on the energy difference between modified and unmodified shooting point. If accepted, the equations of motion are integrated forward and backward in time until a certain path length is obtained or until the condition function $\hat{h}(\mathbf{x})$ can be assigned 0 or 1. In the last case the trial move will be accepted. Any rejection along this scheme implies that the whole trial path will be rejected and the old path is counted again just like in standard Metropolis MC. Naturally, the random walk in path space should obey detailed balance

$$\frac{P_{\text{gen}}[\mathbf{x}^{(\text{o})} \to \mathbf{x}^{(\text{n})}]}{P_{\text{gen}}[\mathbf{x}^{(\text{n})} \to \mathbf{x}^{(\text{o})}]} \frac{P_{\text{acc}}[\mathbf{x}^{(\text{o})} \to \mathbf{x}^{(\text{n})}]}{P_{\text{acc}}[\mathbf{x}^{(\text{n})} \to \mathbf{x}^{(\text{o})}]} = \frac{P[\mathbf{x}^{(\text{n})}]}{P[\mathbf{x}^{(\text{o})}]}, \tag{14}$$

where the superscripts (o) and (n) denote the old and new path, respectively, and $P_{\text{gen}}[\mathbf{x} \to \mathbf{x}']$ is the probability to generate path $\mathbf{x}'$ staring from $\mathbf{x}$. Following the Metropolis-Hastings scheme, the acceptance rule of the whole shooting move can be written as

$$P_{\text{acc}}[\mathbf{x}^{(\text{o})} \to \mathbf{x}^{(\text{n})}] = \hat{h}(\mathbf{x}^{(\text{n})}) \min \left[1, \frac{P[\mathbf{x}^{(\text{n})}]}{P[\mathbf{x}^{(\text{o})}]} \frac{P_{\text{gen}}[\mathbf{x}^{(\text{n})} \to \mathbf{x}^{(\text{o})}]}{P_{\text{gen}}[\mathbf{x}^{(\text{o})} \to \mathbf{x}^{(\text{n})}]} \right] \tag{15}$$

The generation probability is a product of different subprobabilities. These are P_{sel}, to select the shooting point, P_{ran}, for the random modification of this shooting point, and P_{traj}, which is the probability to obtain $\mathbf{x}^{(\text{n})}$ by integrating the equations of motion backward and forward in time starting from the modified shooting point.

$$P_{\text{gen}}[\mathbf{x} \to \mathbf{x}'] = P_{\text{sel}}(x_{\text{shoot}}|\mathbf{x})\, P_{\text{ran}}(x_{\text{shoot}} \to x'_{\text{shoot}})\, P_{\text{traj}}(\mathbf{x}'|x'_{\text{shoot}}) \tag{16}$$

If we generate paths of a fixed length n and if each timeslice has an equal probability to be selected then $P_{\text{sel}} = 1/n$. We come back to this point later on. In addition, TPS algorithms generally utilize a symmetric random modification of the shooting point: $P_{\text{ran}}(x_{\text{shoot}} \to x'_{\text{shoot}}) = P_{\text{ran}}(x'_{\text{shoot}} \to x_{\text{shoot}})$. Therefore, both P_{sel} and P_{ran} cancel out in Eq. (15). The acceptance rule simplifies even further if we also assume that the dynamics obey the microscopic reversibility condition

$$\rho(x)p(x \to y) = \rho(y)p(\bar{y} \to \bar{x}) \tag{17}$$

where $\bar{x}$ is the phasepoint x with reversed velocities: $\bar{x} = (r, -v)$. This relation is very general and valid for a broad class of dynamics applying to both equilibrium and nonequilibrium systems [7]. By applying Eq. (17) several times on Eq. (13) we can show that

$$\begin{aligned} P[\mathbf{x}] = {} & \rho(x_{jdt})p(\bar{x}_{jdt} \to \bar{x}_{(j-1)dt})p(\bar{x}_{(j-1)dt} \to \bar{x}_{(j-2)dt}) \cdots p(\bar{x}_{dt} \to \bar{x}_0) \\ & \times p(x_{jdt} \to x_{(j+1)dt})p(x_{(j+1)dt} \to x_{(j+2)dt}) \cdots p(x_{(n-1)dt} \to x_{ndt}) \end{aligned} \tag{18}$$

is true for any timeslice j. For time-reversible dynamics the backward integration is simply obtained by reversing the velocities and integration forward in time. Hence, the generation probability $P_{\text{traj}}(\mathbf{x}|x_{\text{shoot}})$ depends on exactly the same transition probabilities $p(x \to x')$ and $p(\bar{x} \to \bar{x}')$. This implies that all terms cancel out except the phasepoint density of the shooting point

$$P_{\text{acc}}[\mathbf{x}^{(\text{o})} \to \mathbf{x}^{(\text{n})}] = \hat{h}(\mathbf{x}^{(\text{n})}) \min\left[1, \frac{\rho\left(x^{(n)}_{\text{shoot}}\right)}{\rho\left(x^{(\text{o})}_{\text{shoot}}\right)}\right] \tag{19}$$

This is very convenient as this acceptance/rejection step can take place before the expensive trajectory generation takes place. Still, some (partly) completed trajectories will be rejected in the end due to the condition $\hat{h}(\mathbf{x})$. However, as the new trajectory was generated from a small modification of an existing trajectory with $\hat{h}(\mathbf{x}) = 1$ the chances are relatively high that the condition will be satisfied for the trial trajectory as well.

The sampling of trajectories under a given condition $\hat{h}$ might benefit from using a path ensemble that has a nonfixed length. Using a fixed path length to sample all possible trajectories between A and B is expensive as this length needs to be adapted to the longest pathway connecting these states. Many trajectories will

reach A to B in a much shorter time and will, therefore, consist of unnecessary parts that are not relevant for the actual barrier crossing event. In addition, if trajectories have significant parts outside the barrier region, the shooting move becomes inefficient as many shooting points will lie inside the reactant or product well. Shooting from these points gives a very low probability to connect both states. Using flexible path lengths was first introduced in Ref. 8 within the context of the TIS rate evaluation. However, also for the generation of reactive trajectories, the flexible path ensemble is very useful and allows to generate paths that start and end just at the boundaries of A and B (see Fig. 2). The only difference with the previous example is that $P_{\text{sel}} = 1/n$ is not canceled as the trajectory length can be different. Therefore, if the shooting procedure selects the timeslices by an equal probability, the acceptance rule becomes

$$P_{\text{acc}}[\mathbf{x}^{(\text{o})} \to \mathbf{x}^{(\text{n})}] = \hat{h}(\mathbf{x}^{(\text{n})}) \min \left[1, \frac{\rho\left(x^{(\text{n})}_{\text{shoot}}\right)}{\rho\left(x^{(\text{o})}_{\text{shoot}}\right)} \frac{n^{(\text{o})}}{n^{(\text{n})}} \right] \tag{20}$$

with $n^{(\text{o})}, n^{(\text{n})}$ the length of the old and new path. This expression is not so convenient as a rejection can only be made whenever the whole path is completed. Hence, the integration needs to be carried out even if $\rho(x^{(\text{n})}_{\text{shoot}}) \ll \rho(x^{(\text{o})}_{\text{shoot}})$ implying an almost certain rejection. We can, however, separate the acceptance into two steps by writing

$$P_{\text{acc}}[\mathbf{x}^{(\text{o})} \to \mathbf{x}^{(\text{n})}] = \hat{h}(\mathbf{x}^{(\text{n})}) \min \left[1, \frac{\rho\left(x^{(\text{n})}_{\text{shoot}}\right)}{\rho\left(x^{(\text{o})}_{\text{shoot}}\right)} \right] \times \min \left[1, \frac{n^{(\text{o})}}{n^{(\text{n})}} \right] \tag{21}$$

This acceptance rule obeys detailed balance as well and allows to reject a modification of the shooting move that gives a too high energy. Still, even if the first step is accepted, the final trajectory might be rejected whenever it becomes too long compared to the previous path. We can improve the efficiency even further using following trick [8]. Instead of taking a random number $\alpha \in [0:1]$ after finishing our trajectory and then accept if $\alpha < n^{(\text{o})}/n^{(\text{n})}$, we will actually draw this random number before starting the integration of motions. As we now know that we will have to reject our trajectory whenever $\alpha < n^{(\text{o})}/n^{(\text{n})}$, we can simply define a maximum allowed path length of this trial move in advance

$$n^{\max} = \text{int}[n^{(\text{o})}/\alpha] \tag{22}$$

This allows to directly stop our trial move whenever it exceeds this maximum path length.

The original TPS method also provided an algorithm to determine the reaction rate of the process. This approach has been improved by the TIS [8] and RETIS [9, 10] algorithms. Like RF, the TPS rate evaluation does require a RC (I will not

make the distinction between orderparameter or RC). However, one can show that, compared to the RF method, the efficiency of TPS, TIS, and RETIS, is less sensitive to an improper choice of the RC [39].

The original TPS rate evaluation is based on following correlation function

$$C(t) = \frac{\langle h_A(x_0) h_B(x_t) \rangle}{\langle h_A(x_0) \rangle} \tag{23}$$

where $h_{A/B}(x) = 1$ if $x \in A/B$ and 0 otherwise. Just like Eq. (11), $C(t)$ will initially show some oscillations. However, if there is a separation of timescales, this correlation function grows linearly in time, $C(t) \sim k_{AB} t$, for times $\tau_{\text{mol}} < t < \tau_{\text{rxn}}$. Hence,

$$\tilde{k}_{AB}(t) = \frac{d}{dt} C(t) = \frac{\langle h_A(x_0) \dot{h}_B(x_t) \rangle}{\langle h_A(x_0) \rangle}, \quad k_{AB} = \tilde{k}_{AB}(t') \text{ for } \tau_{\text{mol}} < t' < \tau_{\text{rxn}} \tag{24}$$

The correlation function $C(t)$ is calculated in the TPS scheme using the shooting algorithm in combination with umbrella sampling. First, the fixed path length t' is fixed to a value where $C(t)$ should give a plateau. Then a series of path sampling simulations will be performed in which the final region B is slowly shrunk in successive steps from the entire phase space to the final stable state B [7]. For each step numerous trajectories are generated with that condition that the path should start in A and end in the extended region B at time t'. The distribution of the path's endpoint will be binned into histograms that will be matched just like ordinary umbrella sampling. Once the fully matched histogram is obtained, $C(t)$ is obtained by integration of this histogram over the actual region B.

The approach is rather time-consuming because it can take a relatively long time τ_{mol} before $C(t)$ reaches a plateau (longer than in a transmission coefficient calculation [7]). In Ref. 5 an improvement of this approach was presented in which the umbrella sampling series could be performed with paths shorter than τ_{mol}. The results were then corrected by a factor that is obtained from a single path sampling simulation using the longer paths. Unfortunately, the relative error in this correction factor is large if the path length is reduced too much, so that the gain in CPU efficiency remains limited [8]. Moreover, inspection of Eqs. (23) and (24) shows that a necessary cancellation of positive and negative terms can slow down the convergence of the MC sampling procedure.

IV. TRANSITION INTERFACE SAMPLING

TIS is a more efficient way to calculate reaction constant than the method discussed above. The TIS methodology is also the basis of several other algorithms [9, 11, 13, 44, 45] of which the PPTIS, RETIS, and FFS methods will be discussed in following sections. The TIS rate equation is related to both the EPF expression,

Eq. (9), and to the correlation function used in TPS, Eq. (23), albeit using different kind of characteristic functions. Instead of using the characteristic functions of the stable states A and B, we will redefine the correlation function using *overall states* $\mathcal{A}$ and $\mathcal{B}$. These states do not only depend on the position at the time of consideration but also on its past behavior. Overall state $\mathcal{A}$ covers all phase space points lying inside stable region A, which constitutes the largest part, but also all phase space points that visit A, before reaching B when the equations of motion are integrated backward in time. In other words, all phasepoints that were more recently in A rather than in B. Similarly, state $\mathcal{B}$ comprises stable state B and all phasepoints, coming directly from this state in the past, that is, without having been in A. The corresponding correlation function is

$$C(t) = \frac{\langle h_{\mathcal{A}}(x_0) h_{\mathcal{B}}(x_t)\rangle}{\langle h_{\mathcal{A}}(x_0)\rangle} \tag{25}$$

where $h_{\mathcal{A}}(x_0) = h^b_{AB}(x_0)$ and $h_{\mathcal{B}}(x_0) = h^b_{BA}(x_0)$. Contrary to Eq. (23), this correlation function has no oscillatory behavior during a molecular timescale τ_{mol}. On the contrary, it exhibits a linear regime $\sim k_{AB}t$ for $0 < t < \tau_{\mathrm{rxn}}$. The system will only transfer from overall state $\mathcal{A}$ to overall state $\mathcal{B}$ when it enters region B for the first time since it left region A. If it leaves state B shortly thereafter, it will remain in $\mathcal{B}$. Therefore, $h_{\mathcal{B}}(x_t)$ and $h_{\mathcal{A}}(x_t)$ do not show the fast fluctuations that are found for $h_B(x_t)$ and $h_A(x_t)$. As Eq. (25) is linear from the start, we can simply take the time derivative at $t = 0$, which gives

$$\begin{aligned} k_{AB} &= \frac{\left\langle h_{\mathcal{A}}(x_0)\dot{h}_{\mathcal{B}}(x_0)\right\rangle}{\langle h_{\mathcal{A}}(x_0)\rangle} = \lim_{dt\to 0}\frac{1}{dt}\frac{\left\langle h^b_{AB}(x_0)\theta(\lambda_B - \lambda(x_0))\theta(\lambda(x_{dt}) - \lambda_B)\right\rangle}{\langle h_{\mathcal{A}}(x_0)\rangle} \\ &= \frac{\left\langle h^b_{AB}(x_0)\dot{\lambda}\,(x_0)\,\delta(\lambda(x_0) - \lambda_B)\theta\left(\dot{\lambda}\,(x_0)\right)\right\rangle}{\langle h_{\mathcal{A}}(x_0)\rangle} \end{aligned} \tag{26}$$

The resulting expression is basically the EPF expression [Eq. (9)] through the interface λ_B. At a first sight it might seem that generating long trajectories is no longer needed. As we only need $dC(t)/dt$ at $t = 0$, the minimum time range over which we need to calculate $C(t)$ is $[0 : dt]$ instead of $[0 : \tau_{\mathrm{mol}}]$. Unfortunately, unlike $h_A(x_0)$ and $h_B(x_0)$, the determination of $h_{\mathcal{A}}(x_0)$ and $h_{\mathcal{B}}(x_0)$ cannot be done instantaneously. For this we still need to integrate the equations of motion. However, for most x_0, $h_{\mathcal{A}}(x_0)/h_{\mathcal{B}}(x_0)$ can be assigned 1 or 0 using a much shorter backward trajectory than τ_{mol}. For stochastic dynamics $h_{\mathcal{A}}(x_0)/h_{\mathcal{B}}(x_0)$ can be, strictly speaking, a fractional number. However, there is generally no need to know this fractional number for a specific phasepoint, except for committor analysis [14, 46–49]. Hence, TIS algorithms will generally compute $h_A(x_0)/h_B(x_0)$ for one specific path to which x_0 belongs. Conceptually, it is therefore more accurate to speak of a MC sampling in pathspace rather than phase space. The TIS correlation function

has an additional advantage that the reaction rate is somewhat better defined if the separation of timescales $\tau_{\rm mol} \ll \tau_{\rm rxn}$ is not sufficiently obeyed. The fact that the derivative of $C(t)$ is taken at $t = 0$ makes corrections like the one suggested in Ref. 50 unnecessary.

The TIS algorithm expresses the rate equation, Eq. (26), as a product of different terms. Each term has a much higher value than the final rate and is, therefore, much easier to compute. To introduce the TIS and PPTIS expression, that I will discuss in the next section, it is convenient to introduce following crossing probabilities that depend on four nonintersecting interfaces $\{x|\lambda(x) = \lambda_i\}$, $\{x|\lambda(x) = \lambda_j\}$, $\{x|\lambda(x) = \lambda_k\}$, $\{x|\lambda(x) = \lambda_l\}$

$$P\left(\begin{smallmatrix} k \\ l \end{smallmatrix}\middle|\begin{smallmatrix} j \\ i \end{smallmatrix}\right) = \lim_{dt\to 0} \frac{\langle h^b_{ij}(x_0)\theta(\lambda_j - \lambda(x_0))\theta(\lambda(x_{dt}) - \lambda_j)h^f_{kl}(x_0)\rangle}{\langle h^b_{ij}(x_0)\theta(\lambda_j - \lambda(x_0))\theta(\lambda(x_{dt}) - \lambda_j)\rangle}$$
$$= \frac{\langle h^b_{ij}(x_0)\dot\lambda(x_0)\delta(\lambda(x_0) - \lambda_j)\theta(\dot\lambda(x_0))h^f_{kl}(x_0)\rangle}{\langle h^b_{ij}(x_0)\dot\lambda(x_0)\delta(\lambda(x_0) - \lambda_j)\theta(\dot\lambda(x_0))\rangle} \text{ for } \lambda_j > \lambda_i \quad (27)$$

For $\lambda_j < \lambda_i$, we simply need to replace $\theta(\lambda_j - \lambda(x_0))\theta(\lambda(x_{dt}) - \lambda_j)$ by $\theta(\lambda(x_0) - \lambda_j)\theta(\lambda_j - \lambda(x_{dt}))$ in the first line or $\dot\lambda(x_0)\delta(\lambda(x_0) - \lambda_j)\theta(\dot\lambda(x_0))$ by $-\dot\lambda(x_0)\delta(\lambda(x_0) - \lambda_j)\theta(-\dot\lambda(x_0))$ in the second line of the above definition. Equation (27) defines a conditional crossing probability. It is the probability that the system will cross interface λ_k before λ_l under a twofold condition. These conditions are that the system should cross interface λ_j right now at time $t = 0$, while λ_i was more recently crossed than λ_j in the past.

Using these crossing probabilities, one can prove that Eq. (26) is equivalent to the product of the initial flux times the overall crossing probability [8]

$$k_{AB} = \frac{\langle\dot\lambda(x_0)\delta(\lambda(x_0) - \lambda_0)\theta(\dot\lambda(x_0))\rangle}{\langle h_{\mathcal{A}}(x_0)\rangle} \times P(^n_0|^0_{0^-}) \equiv f_A\mathcal{P}_A(\lambda_n|\lambda_0) \quad (28)$$

where λ_0 and λ_n are the boundaries of the stable states A and B. f_A is just the flux out of state A that can be computed with standard MD as the boundary of A is set at the left side of the barrier region. The minus in 0^- is to denote an interface $\lambda_0 - \epsilon$ that is put there to indicate the direction of the crossing at $t = 0$. The overall crossing probability $\mathcal{P}_A(\lambda_n|\lambda_0) = P(^n_0|^0_{0^-})$ is the probability that once λ_0 is crossed, λ_n will be crossed before a recrossing with λ_0 occurs. This probability is very small, but it can be calculated by defining $n - 1$ nonintersecting interfaces in between λ_0 and λ_n and express the overall crossing probability as the following product [8]

$$\mathcal{P}_A(\lambda_n|\lambda_0) = \prod_{i=0}^{n-1} \mathcal{P}_A(\lambda_{i+1}|\lambda_i) \quad (29)$$

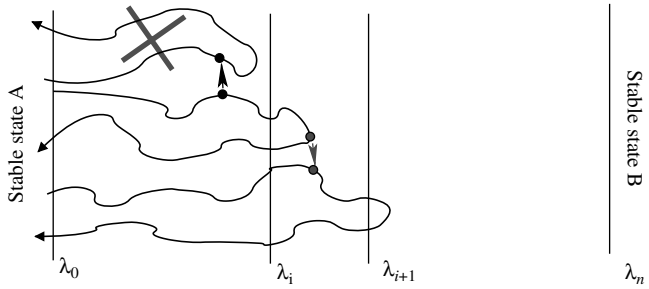


Figure 3. The TIS path ensemble $[i^+]$ is required to calculate the conditional crossing probability $\mathcal{P}_A(\lambda_{i+1}|\lambda_i)$. For this purpose we apply the shooting move to generate all possible trajectories starting at λ_0 and ending at λ_0 or λ_n with at least one crossing with λ_i. Suppose the algorithm starts with the middle path that already fulfills these requirements. A shooting point is randomly selected and modified. However, the trajectory that starts out from this point (top trajectory) fails to cross λ_i and is therefore rejected and the old path is counted again. A new shooting generates a valid trajectory that not only crosses λ_i but λ_{i+1} as well. This trajectory is called "successful." The fraction of successful trajectories in this ensemble determine $\mathcal{P}_A(\lambda_{i+1}|\lambda_i)$.

The factorization of $\mathcal{P}_A(\lambda_n|\lambda_0)$ into probabilities $\mathcal{P}_A(\lambda_{i+1}|\lambda_i)$ that are much higher than the overall crossing probability, is the basis of the importance sampling approach. It is important to note that $\mathcal{P}_A(\lambda_{i+1}|\lambda_i)$ are in fact complicated history dependent conditional probabilities. If we consider all possible pathways that start at λ_A and end by either crossing λ_A or λ_B, while have at least one crossing with λ_i in between, the fraction that crosses λ_{i+1} as well equals $\mathcal{P}_A(\lambda_{i+1}|\lambda_i)$. This basically reduces the problem to a correct sampling of trajectories that should obey the λ_i crossing condition (see Fig. 3). From now on we will call this the $[i^+]$ path ensemble. In TIS, this is done via the shooting algorithm for flexible path lengths as is discussed in previous section. (For a full flowchart diagram of the TIS algorithm, see Ref. 51.) The number of interfaces and their separation should be set to maximize efficiency. In Refs 39, 40, it was found that the optimal interface separation is obtained when one out of five trajectories reach the next interface. In addition, one can define a set of subinterfaces of arbitrary separation in order to construct the crossing probability as a continuous function. This strictly decreasing function could be viewed as the dynamical analogue of the free energy profile $F(\lambda)$.

There are some small differences how to treat the path ensemble $[i^+]$ regarding the end point of the path. In the first TIS algorithms, the trajectory could reach up to λ_{i+1}, where the trajectory was stopped and assigned successful. More recent simulations continue the trajectory until reaching the stable states A or B each time. The additional cost is very limited as about 80% is not reaching λ_{i+1} and need to be followed until reaching A anyway. The choice to continue the trajectory even

after λ_{i+1} has the advantage that one can start the $[i^+]$ path ensemble without the need to fix a value for λ_{i+1} beforehand. After some simulation cycles the λ_{i+1} can be set to have the optimal 20% success rate after which one can start the $[(i+1)^+]$ path ensemble. In addition, the new approach makes it much more easy to use replica exchange which we will discuss in Section VII.

The simplicity of Eq. (29) is deceptive and could be mistaken as a Markovian approximation. The reason that the equation is still exact lies in the fact that the crossing probabilities are history dependent and by the fact that it only considers first crossing events. We can argue the exactness of the equations also in another way. Suppose we want to calculate the probability to go from λ_0 to λ_1 to $\lambda_2 \ldots$ to λ_n in successive jumps. This probability can be expressed as

$$\begin{aligned} &P(\lambda_0 \to \lambda_1 \to \lambda_2 \to \cdots \to \lambda_n) \\ &\quad = P(\lambda_0) \times P(\lambda_0 \to \lambda_1 | \lambda_0) \times P(\lambda_1 \to \lambda_2 | \lambda_0 \to \lambda_1) \\ &\qquad \times P(\lambda_2 \to \lambda_3 | \lambda_0 \to \lambda_1 \to \lambda_2) \times \cdots \\ &\qquad \times P(\lambda_{n-1} \to \lambda_n | \lambda_0 \to \lambda_1 \to \lambda_2 \to \cdots \lambda_{n-2} \to \lambda_{n-1}) \end{aligned} \tag{30}$$

This is an exact non-Markovian expression for this specific crossing sequence that looks similar to Eq. (29). However, it does not say anything about the many different trajectories that could connect λ_0 with λ_n. For instance, we should also take into account the sequence $\lambda_0 \to \lambda_1 \to \lambda_2 \to \lambda_1 \to \lambda_2 \to \lambda_3 \to \ldots \lambda_{n-1} \to \lambda_n$. Therefore, it might seem that the right expression should look much more complicated than Eq. (29). The trick, however, is that this last sequence cannot occur if we only consider first crossing events. When we move back to λ_1 in the third step, this move will simply not considered as it is a second visit since leaving λ_0. Hence, the successive sequence $\lambda_0 \to \lambda_1 \to \lambda_2 \to \cdots \lambda_{n-1} \to \lambda_n$ is the only possible sequence of first crossing events that brings you from λ_0 to λ_n.

V. PARTIAL PATH SAMPLING

The PPTIS is a variation of the TIS algorithm that was devised to treat diffusive barrier crossings [11]. Despite the existence of a fine separation of timescale, that is, the time to cross the barrier is still negligible compared to the time spend in the reactant well, the path length can become too long for an effective computation of the reaction rate. This is the case if the barriers are sufficiently high to ensure exponential relaxation, but not very sharp so that system can move backward and forward on the barrier before it eventually drops off. The PPTIS equation depends on the same rate equation as TIS

$$k_{AB} = f_A P \left({n \atop 0} \middle| {1 \atop 0} \right) \tag{31}$$

The only difference with Eq. (28) is that we now consider the condition $|_0^1)$ instead of $|_{0^-}^0)$, but this is just a technical detail. If we take $\lambda_1 = \lambda_0 + \epsilon$ the equations become equivalent. However, λ_1 can be any value that is somewhat larger than λ_0 by redefining f_A as the effective flux through λ_1. This implies that we should count the positive crossings with λ_1 whenever the system leaves the stable state $A : \{x|\lambda(x) < \lambda_0\}$. However, the next positive crossing should only be counted if the system has revisited A again. The PPTIS approach tries to avoid the generation of very long trajectories using a soft Markovian approximation. The PPTIS scheme assumes that for a well positioned set of interfaces the system will lose its memory over a distance that is similar to the interface separation. This implies for any $m > 1$

$$P\left(\begin{smallmatrix} l \\ k \end{smallmatrix}|^j_{j\pm m}\right) \approx P\left(\begin{smallmatrix} l \\ k \end{smallmatrix}|^j_{j\pm 1}\right) \tag{32}$$

The PPTIS algorithm consist again of a series of path sampling simulations. Each PPTIS simulation samples a certain path ensemble in which trajectories are confined within two next-nearest interfaces. For instance, the $[i^{\pm}]$ path ensemble will consist of all possible trajectories starting and ending at either λ_{i+1} or λ_{i-1} having at least one crossing with the middle interface λ_i. From these simulations the following short-distance crossing can be obtained

$$p_i^{\pm} \equiv P\left(\begin{smallmatrix} i+1 \\ i-1 \end{smallmatrix}|^i_{i-1}\right), p_i^{=} \equiv P\left(\begin{smallmatrix} i-1 \\ i+1 \end{smallmatrix}|^i_{i-1}\right), p_i^{\mp} \equiv P\left(\begin{smallmatrix} i-1 \\ i+1 \end{smallmatrix}|^i_{i+1}\right), p_i^{\ddagger} \equiv P\left(\begin{smallmatrix} i+1 \\ i-1 \end{smallmatrix}|^i_{i+1}\right) \tag{33}$$

with $p_i^{\pm} + p_i^{=} = p_i^{\mp} + p_i^{\ddagger} = 1$. For instance, $p_i^{\pm}$ is determined by dividing the number of trajectories in the $[i^{\pm}]$ ensemble that start at λ_{i-1} and end at λ_{i+1} divided by all trajectories that start at λ_{i-1}.

Once these short-distance crossing probabilities are obtained with sufficient accuracy, the overall crossing probability can be obtained. One way to do this is to use these probabilities as input for a kinetic MC simulation [52]. However, this is not needed for a one-dimensional RC that allows an elegant analytical treatment [11]. Naturally, $P(\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}|^1_0) = p_1^{\pm}$, but the calculation of $P(\begin{smallmatrix} 3 \\ 0 \end{smallmatrix}|^1_0)$ requires already to sum up the trajectories $0 \to 1 \to 2 \to 3$, $0 \to 1 \to 2 \to 1 \to 2 \to 3$, and so on. However, as shown in Ref. 11, one can derive following recursive relations to make this infinite summation of all trajectories (include the ones of infinite length!). These PPTIS recursive relations are the following

$$P\left(\begin{smallmatrix} m+1 \\ 0 \end{smallmatrix}|^1_0\right) = \frac{p_m^{\pm} P\left(\begin{smallmatrix} m \\ 0 \end{smallmatrix}|^1_0\right)}{p_m^{\pm} + p_m^{=} P\left(\begin{smallmatrix} 0 \\ m \end{smallmatrix}|^{m-1}_m\right)}, \quad P\left(\begin{smallmatrix} 0 \\ m+1 \end{smallmatrix}|^m_{m+1}\right) = \frac{p_m^{\mp} P\left(\begin{smallmatrix} 0 \\ m \end{smallmatrix}|^{m-1}_m\right)}{p_m^{\pm} + p_m^{=} P\left(\begin{smallmatrix} 0 \\ m \end{smallmatrix}|^{m-1}_m\right)} \tag{34}$$

or, by defining the long-distance crossing probabilities $P_m^+ \equiv P\left(\begin{smallmatrix} m \\ 0 \end{smallmatrix}\middle|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, $P_m^- = P\left(\begin{smallmatrix} 0 \\ m \end{smallmatrix}\middle|\begin{smallmatrix} m-1 \\ m \end{smallmatrix}\right)$

$$k_{AB} = f_A P_n^+$$
$$P_{m+1}^+ = \frac{p_m^{\pm} P_m^+}{p_m^{\pm} + p_m^{=} P_m^-}, \quad P_{m+1}^- = \frac{p_m^{\mp} P_m^-}{p_m^{\pm} + p_m^{=} P_m^-}, \quad \text{for } m > 1, \quad P_1^+ = P_1^- = 1 \tag{35}$$

Hence, starting from the initial conditions for $(P_1^+, P_1^-) = (1, 1)$, one can successively solve $(P_2^+, P_2^-), (P_3^+, P_3^-), \ldots, (P_n^+, P_n^-)$ via Eq. (35). It is important to note that P_j^- is not exactly the same as P_j^+ in the reverse direction. Only for $j = n$ these two probabilities can be viewed as mirror images.

Here, I will derive an alternative recursive relation that does not require the auxiliary reverse probabilities P_j^-. The derivation is similar to the one presented in the supplemental information of Ref. 53 that treats the simpler hopping process. To achieve this, we will bring $P(\begin{smallmatrix} 0 \\ m \end{smallmatrix}|\begin{smallmatrix} m-1 \\ m \end{smallmatrix})$ in front of Eq. (34).

$$P\left(\begin{smallmatrix} 0 \\ m \end{smallmatrix}\middle|\begin{smallmatrix} m-1 \\ m \end{smallmatrix}\right) = \frac{p_m^{\pm}\left[P\left(\begin{smallmatrix} m \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) - P\left(\begin{smallmatrix} m+1 \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)\right]}{p_m^{=} P\left(\begin{smallmatrix} m \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)} \tag{36}$$

or by incrementing m

$$P\left(\begin{smallmatrix} 0 \\ m+1 \end{smallmatrix}\middle|\begin{smallmatrix} m \\ m+1 \end{smallmatrix}\right) = \frac{p_{m+1}^{\pm}\left[P\left(\begin{smallmatrix} m+1 \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) - P\left(\begin{smallmatrix} m+2 \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)\right]}{p_{m+1}^{=} P\left(\begin{smallmatrix} m+2 \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)} \tag{37}$$

Moreover, we can write for $P(\begin{smallmatrix} m+1 \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix})$

$$\begin{aligned} P\left(\begin{smallmatrix} m+1 \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) &= P\left(\begin{smallmatrix} m \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) P\left(\begin{smallmatrix} m+1 \\ 0 \end{smallmatrix}|\begin{smallmatrix} m \\ m-1 \end{smallmatrix}\right) \\ &= P\left(\begin{smallmatrix} m \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)\left(1 - P\left(\begin{smallmatrix} 0 \\ m+1 \end{smallmatrix}|\begin{smallmatrix} m \\ m-1 \end{smallmatrix}\right)\right) \\ &= P\left(\begin{smallmatrix} m \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)\left(1 - P\left(\begin{smallmatrix} 0 \\ m+1 \end{smallmatrix}|\begin{smallmatrix} m \\ m+1 \end{smallmatrix}\right) p_m^{=}/p_m^{\mp}\right) \end{aligned} \tag{38}$$

Then, we substitute Eq. (37) in Eq. (38) and bring $P(\begin{smallmatrix} m+2 \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix})$ in front which yields

$$P\left(\begin{smallmatrix} m+2 \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) = \frac{p_m^{=} p_{m+1}^{\pm} P\left(\begin{smallmatrix} m \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) P\left(\begin{smallmatrix} m+1 \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)}{\left(p_m^{\mp} p_{m+1}^{=} + p_m^{=} p_{m+1}^{\pm}\right) P\left(\begin{smallmatrix} m \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) - p_m^{\mp} p_{m+1}^{=} P\left(\begin{smallmatrix} m+1 \\ 0 \end{smallmatrix}|\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)} \tag{39}$$

or

$$k_{AB} = f_A P_n^+$$
$$P_{m+2}^+ = \frac{p_m^{=} p_{m+1}^{\pm} P_m^+ P_{m+1}^+}{\left(p_m^{\mp} p_{m+1}^{=} + p_m^{=} p_{m+1}^{\pm}\right) P_m^+ - p_m^{\mp} p_{m+1}^{=} P_{m+1}^+}, \quad P_1^+ = 1, \quad P_2^+ = p_1^{\pm} \tag{40}$$

In the case of a full Markovian assumption, $p_i^{\pm} = p_i^{\dagger}$ and $p_i^{\mp} = p_i^{=}$, we reobtain the simpler expression of Ref. 53. For some this new expression might seem more esthetic as the set of two equations has now been transferred into a single recursive equation without relying on the auxiliary probability P_m^-. The new function has its utility [53], but is numerically somewhat problematic as it can sometimes produce zeros in both nominator and denominator that cancel, but is not practical for numerical calculations.

The positioning of interfaces is crucial in PPTIS. On one hand, one would like to put them close together to improve efficiency. On the other hand, putting them to close will introduce systematic errors due to a decrease of history dependence of the hopping probabilities, which invalidates Eq. (32). A way to measure whether the interfaces are sufficiently far is the calculation of a memory-loss function [11]. However, the memory-loss function can only provide a necessary but not necessarily sufficient condition for this separation. In Fig. 4 we give two examples of well and badly placed interfaces.

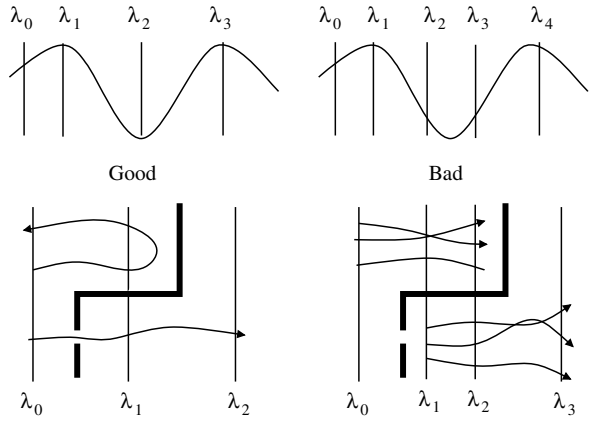


Figure 4. Examples of well and badly positioned interfaces with respect to the memory-loss assumption Eq. (32). The top situation requires a good description of the kinetic correlations whenever the top of the barrier has many small local wells. At the left, the interfaces are correctly placed. Once the system crosses λ_1 it will gain a lot of kinetic energy when arriving at λ_2. Henceforth, the chances are high that the system would not get trapped and directly moves upward to cross λ_3. The PPTIS simulation for this interface configuration will show that $p_2^{\pm} \gg \frac{1}{2}$ as it should. At the right-hand side we have put an additional interface inside the local well. The $[3^{\pm}]$ path ensemble will consists of trajectories having a much lower kinetic energy than the $[2^{\pm}]$ ensemble of the left-hand side. Henceforth, the right-hand side will overestimate the probability to get trapped. The bottom picture shows impermeable wall (thick black line) with a small hole. The left-hand side shows a correct positioning of interfaces. The pathways that are generated from λ_0 to λ_2 all have to move through the small hole. Conversely, the right-hand side shows a bad overlap between the $[1^{\pm}]$ and $[2^{\pm}]$ path ensembles which might give the impression that trajectories can tunnel through the wall.

The milestoning method [12] is very similar to PPTIS. There are basically two important differences. Milestoning assumes full memory loss once the system hits an interface. In our notation this could be rephrased as $P(^{l}_{k}|^{j}_{j\pm m}) \approx P(^{l}_{k}|^{j}_{j})$. This is a stronger approximation than Eq. (32). The approximation of milestoning becomes exact, if the interfaces coincide with the isocommittor functions [54], but these are difficult determine. On the other hand, milestoning is more precise in the construction of the time evolution of the system by making the crossing probabilities time dependent. This is important if there is not a clear separation of timescales and also allows to calculate other dynamical properties like diffusion. Hence, instead of $p_i^{\pm}$, $p_i^{=}$, $p_i^{\mp}$, $p_i^{\ddagger}$ with $p_i^{\pm} + p_i^{=} = p_i^{\mp} + p_i^{\ddagger} = 1$, milestoning calculates for each interface the time-dependent probability densities $p_i^{+}(t)$, $p_i^{-}(t)$ with $\int_0^{\infty}[p_i^{+}(t) + p_i^{-}(t)]dt = 1$. The strengths of PPTIS and milestoning do not exclude each other and could be unified into a single method as was suggested in Ref. 55. A realization of such a method was recently published [56].

VI. FORWARD FLUX SAMPLING

FFS was originally developed for the special case of biochemical networks that do not obey equilibrium statistics nor time reversibility [13]. However, its advantageous implementation and apparent efficiency has gained this method a fast increasing popularity for equilibrium systems as well. FFS is based on the same theoretical TIS rate equations (28) and (29). However, the fundamental difference is the sampling move. While the principal sampling move in TIS is the shooting move, FFS is based on a non-Metropolis MC scheme called splitting [57, 58]. This approach requires stochastic dynamics, although it has been suggested that FFS is able to treat deterministic dynamics utilizing the Lyapunov instability [59, 60] using small "invisible" stochastic noises [61]. Like TIS, FFS consists of a straightforward MD simulation, from which the escape flux f_A is obtained, followed by a series of path sampling simulations. However, besides giving the flux value, the MD simulation also provides the starting conditions for the path sampling simulations. Each time that the first interface λ_0 is crossed in the positive direction, this phasepoint just after the interface is stored on the hard disk. In the first path simulation, performing the $[0^+]$ path ensemble, these points are used as starting points for the trajectories that are continued until reaching λ_1 or returning to λ_0. Naturally, stochasticity is of eminent importance, otherwise all these trajectories would just reproduce parts of MD simulation. The endpoints of the trajectories that successfully reach λ_1 are stored again and serve as initial points for the $[1^+]$ ensemble. The path ensembles are executed one after the other by the same procedure until reaching state B (see Fig. 5).

There are advantages and disadvantages compared to the TIS algorithm. The most important advantage of FFS is that it allows to treat nonequilibrium systems as it does not require any knowledge about the phasepoint density. TIS employs the

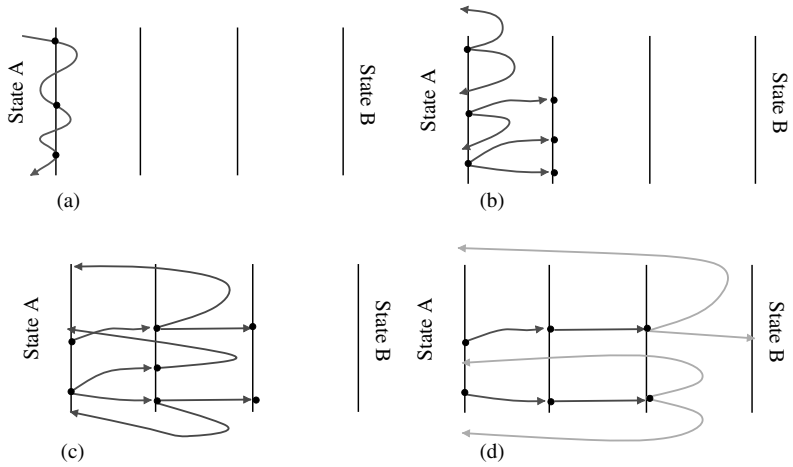


Figure 5. Illustration of FFS algorithm. (a) Shows the initial MD simulation that is needed to calculate the flux f_A. The positive crossing points (black dots) are stored. (b) Starting from the stored MD crossing points a number of trajectories are released. The endpoints of the successful trajectories (that reach λ_1) are stored again and used for the next path ensemble simulation (c). Finally, the reactant state will be reached (d).

shooting move that requires to know $\rho(x)$ for the acceptance, Eq. (19). In addition, FFS does not require any integration of motion backward in time. Therefore, time reversibility is not required. Moreover, unlike PPTIS, FFS is, in principle, equally exact as TIS. However, if one has to chose between TIS and FFS for equilibrium dynamics, one has to consider following points. FFS will generally create more trajectories for the same number of MD steps as it recycles previously generated trajectories. Moreover, there are no rejections like there are in TIS and any other Metropolis based MC scheme. In practice, the reduction in MD steps will be limited to a certain factor ($\approx$2,3) as the unsuccessful trajectories, which is the largest part, have to be followed until reaching state A. On the other hand, the FFS trajectories will be much more correlated than the TIS trajectories. This implies that FFS needs much more trajectories than TIS to obtain the same accuracy. One reason for this, is that FFS generates several trajectories having the same starting point. Absence of stochasticity will result that all these trajectories basically coincide. Fully Brownian motion does not exclude correlation effects either as the successful trajectories starting from the same point will hit the next interface in a confined region. The size of this region is determined by the diffusion orthogonal to the RC and the time it takes to go from interface to the other. Besides correlations within a certain path ensemble, the FFS method also introduces correlations between the different ensembles. This is a crucial difference with TIS where the MD simulation and all path simulations are independent. One of its consequences is that the FFS

is more sensitive to the RC than TIS or even RF [39]. An efficiency analysis of FFS [62] ignores this correlation effect. This can be a rather crude approximation that is probably only valid when interfaces are approximately equal to the isocommittor surfaces. Suppose $\lambda^{\perp}$ denotes a coordinate orthogonal to the RC. Let $\mathcal{P}_A(\lambda_n|\lambda_0;\lambda^{\perp})$ be the overall crossing probability from λ_0 to λ_n starting from a point $\lambda^{\perp}$ on the first interface. Then, the full overall crossing probability is given by

$$\mathcal{P}_A(\lambda_n|\lambda_0) = \int d\lambda^{\perp}\, \mathcal{P}_A(\lambda_n|\lambda_0;\lambda^{\perp})\rho(\lambda^{\perp}|\lambda_0) \tag{41}$$

where $\rho(\lambda^{\perp}|\lambda_0)$ is the probability density of $\lambda^{\perp}$ on interface λ_0. FFS will suffer considerably when the distributions $\rho(\lambda^{\perp}|\lambda_0)$ and $\mathcal{P}_A(\lambda_n|\lambda_0;\lambda^{\perp})$ are not overlapping. In that case, FFS will miss important crossing points that are significant for the rate evaluation. Some studies have shown that FFS can significantly underestimate reaction rates [63, 64] in practical cases. Sampling artifacts like this, are also not yet fully excluded as possible explanation for some surprising results on nonequilibrium nucleation [65].

This issue will be most sensitive to the MD and the first interface ensembles on which all the further results will depend. If $\lambda_1, \lambda_2, \ldots, \lambda_{n-1}$ are isocommittor surfaces then $\mathcal{P}_A(\lambda_n|\lambda_i;\lambda^{\perp})$ is a constant as function of $\lambda^{\perp}$. This eliminates the problem. This is the reason that Borrero et al. devised a FFS scheme in which the interfaces are repositioned on-the-fly in order to obtain a proper RC [66].

TIS has the advantage that it can relax the history of the path via the backward integration. Therefore, the distribution density $\rho(\lambda^{\perp}|\lambda_0)$ can change when considering the different path ensembles. TIS can give correct results even if the sampled distribution of starting points $\lambda^{\perp}$ of the final $[n-1^{+}]$ trajectories do not overlap with the initial MD crossing points [67].

VII. REPLICA EXCHANGE TIS

In Ref. 9, I showed that a special type of replica exchange [68, 69] can significantly improve the TIS efficiency (see also Ref. 10 for some extensions of this approach). A crucial difference with standard RE, which has also been applied to TPS [70], is that the RETIS method does not require additional simulations at elevated temperatures. Instead, swaps are attempted between the different TIS path ensembles. For this purpose, RETIS has replaced the initial MD simulation by another path ensemble, called $[0^{-}]$, that consists of all path that start at $\lambda_0 = \lambda_A$, then go in the opposite direction away from the barrier inside state A, and finally end at λ_0 again. The flux is then obtained from the average path length of the $[0^{-}]$ and $[0^{+}]$ ensembles as follows [9].

$$f_A = \left(\left\langle t_{\text{path}}^{[0^{-}]} \right\rangle + \left\langle t_{\text{path}}^{[0^{+}]} \right\rangle \right)^{-1} \tag{42}$$

where $\langle t_{\text{path}}^{[0^-]} \rangle$, $\langle t_{\text{path}}^{[0^+]} \rangle$ are the average path lengths in the $[0^-]$ and $[0^+]$ path ensembles, respectively.

As the dynamical process is now fully described by path simulations with different interface-crossing conditions, the exchange of trajectories between them becomes extremely efficient, especially if the process possesses multiple reaction channels [67]. The methodology avoids the need of doing additional simulations at elevated temperatures and even gives paths for free as for most swapping moves whole trajectories are being swapped. Only when a swapping between the $[0^-]$ and $[0^+]$ ensembles are attempted, two phase points are interchanged. From the last point of the $[0^-]$ trajectory a new path in the $[0^+]$ is generated. Reversaly, the first point of the old $[0^+]$ path will serve to generate a new path in the $[0^-]$ ensemble by integrating the equations of motion backward in time (see Fig. 6).

The RETIS algorithm is then as follows. At each step it is decided by an equal probability whether a series of shooting or swapping moves will be performed. In

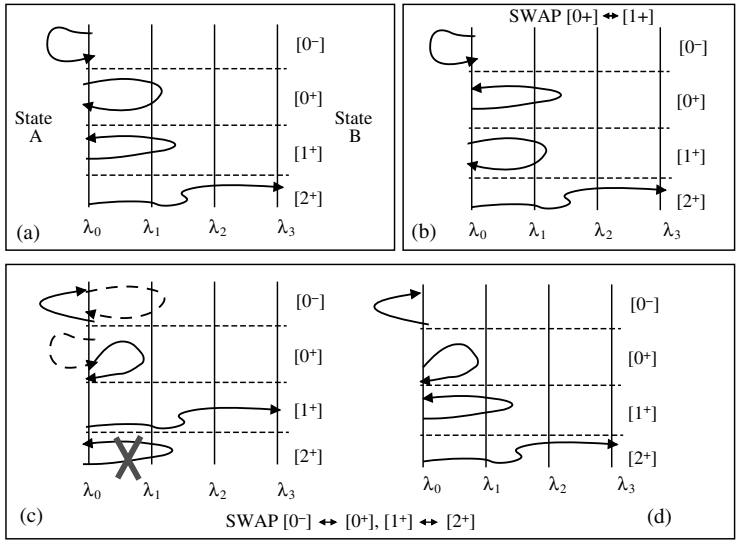


Figure 6. Parallel path swapping move in RETIS. The picture illustrates a RETIS simulation using four interfaces. (a) shows the initial "superstate" that contains one trajectory per ensemble $[0^-]$, $[0^+]$, $[1^+]$, and $[2^+]$. (b) Shows the superstate after the $[0^+] \leftrightarrow [1^+]$ swap. The original $[0^+]$ trajectory crossed λ_1 and is therefore a valid path in the $[1^+]$ ensemble. The swapping move is, therefore, accepted. (c) Shows the trial superstate that is obtained after the simultaneous swaps $[0^-] \leftrightarrow [0^+]$ and $[1^+] \leftrightarrow [2^+]$. The first swap requires the integration of motion forward or backward in time starting from the last or first timeslice of the swapped trajectories. This swap will always generate acceptable trajectories for the $[0^-]$, $[0^+]$ ensembles. The other swap $[1^+] \leftrightarrow [2^+]$ is rejected because the old $[1^+]$ path does not cross λ_2. (d) Gives the final situation after the whole move.

the first case, all simulations will be updated sequentially by one shooting move. In the second case, again an equal probability will decide whether the swaps $[0^-] \leftrightarrow [0^+]$, $[1^+] \leftrightarrow [2^+]$, ... or the swaps $[1^+] \leftrightarrow [2^+]$, $[3^+] \leftrightarrow [4^+]$, ... are performed. Each time that $[0^-]$ and $[(n-1)^+]$ do not participate in the swapping move they are left unchanged. Also when the swapping move does not yield valid paths for both ensembles, the move is rejected for these two simulations and the old paths are counted again. Note that the swapping moves do not require any force calculations except for the $[0^-] \leftrightarrow [0^+]$ swap.

Like FFS, the path ensembles in RETIS are not fully uncorrelated. However, their dependence fundamentally different. In FFS, the path ensemble $[i^+]$ is fully determined by its predecessors, the MD simulation and the path simulations $[j^+]$ with $j < i$. Conversely, the separate RETIS simulations generate a large part of their trajectories independently. The exchange between the ensembles is, therefore, an additional help instead of a strict dependence as it is for FFS. Moreover, the benefit of the exchange works in both directions and is mutual for all ensembles, that is, the $[i^+]$ path ensemble can improve the sampling in both $[(i+1)^+]$ and $[(i-1)^+]$ via the swapping moves $[i^+] \leftrightarrow [(i-1)^+]$ and $[i^+] \leftrightarrow [(i+1)^+]$, and will improve itself due to the same moves.

VIII. NUMERICAL EXAMPLE

We will apply the different methods on a simple one-dimensional test system using Langevin dynamics with finite friction. The Langevin dynamics was chosen because its inhibits stochasticity that is required for FFS. Hitherto, most studies on FFS have applied Brownian dynamics. As the dynamics we are considering are not overdamped, the dimensionality of the system becomes effectively two-dimensional. We could consider the velocity as an orthogonal coordinate [like $\lambda^\perp$ in Eq. (41)]. This makes the choice of a proper RC not such a triviality as one would think at first sight. However, we will simply take the RC to be configurational dependent, which is the standard approach. The system that we will consider consist of a single one-dimensional particle inside a double well potential

$$V(r) = k_4 r^4 - k_2 r^2 \tag{43}$$

with $k_4 = 1$ and $k_2 = 2$. The corresponding potential has a maximum at $r = 0$ and two minima at $r = \pm 1$. We use reduced units where the mass and the Boltzmann constant are set to unity, $k_B = m = 1$. The system is coupled to a Langevin thermostat with friction coefficient $\gamma = 0.3$ and temperature $T = 0.07$. The equations of motion are integrated using MD timestep of $dt = 0.002$. The RF method was applied using the EPF formalism for the transmission coefficient calculation. For this purpose 100,000 trajectories were released from the TST dividing surface $r = 0$. The free energy term $\int_{-\infty}^{0} d\lambda\, e^{-\beta F(\lambda)}$ was obtained by a simple

numerical integration. The RF method is by far the most efficient method for this system because there is basically no error in the free energy calculation and the transmission coefficient is close to unity. The RF results will therefore be the reference for the other methods. The escape flux f_A for TIS, FFS, and PPTIS was determined using a MD simulation of 10,000,000 timesteps. The same MD result was used for these three methods. We performed an additional MD simulation using less timesteps, 4,000,000, for a second FFS calculation in order to see how this effects the final FFS result. I defined eight interfaces $\lambda_0 = -0.9$, $\lambda_1 = -0.8$, $\lambda_2 = -0.7$, $\lambda_3 = -0.6$, $\lambda_4 = -0.5$, $\lambda_5 = -0.4$, $\lambda_6 = -0.3$, and $\lambda_7 = 1.0$. For each path ensemble 20,000 trajectories were generated. For TIS and PPTIS, 50% of the MC moves were shooting moves. I applied the aimless shooting [71] approach in which the velocities at the shooting point are completely regenerated from Maxwellian distribution. However, unlike Ref. 71, shooting points were picked with an equal probability for all timeslices along the path without considering the previous shooting point [9]. The other 50% were time-reversal moves. Time-reversal moves simply change the order of the timeslices of the old path while reversing the velocities. Time reversal can sometimes increase the ergodic sampling and is basically cost-free as it doesnot require any force calculations. However, as aimless shooting is also able to reverse velocities in a single step, the time-reversal move could actually have been omitted for this case. In the RETIS algorithm there was at each step a 25% probability to perform a shooting move, another 25% probability to do a time-reversal move, and a 50% probability to do a replica exchange move. The FFS simulations consist of a single move which is the forward integration of the equations of motion. The Langevin thermostat served for the necessarily stochasticity. The results are shown in Table I. The RF method gives the most accurate results as expected. The value for κ can be

TABLE I
Results of the Rate Evaluations Using RF, TIS, PPTIS, RETIS, and FFS

Reactive Flux Method	$\frac{1}{\sqrt{2\pi\beta m}}$	$\frac{e^{-\beta F(0)}}{\int_{-\infty}^{0} d\lambda\, e^{-\beta F(\lambda)}}$	κ	$k = \kappa \times \frac{1}{\sqrt{2\pi\beta m}} \times \frac{e^{-\beta F(0)}}{\int_{-\infty}^{0} d\lambda\, e^{-\beta F(\lambda)}}$
EPF algorithm	0.106	$2.63 \cdot 10^{-6}$	$0.874 \pm 4\%$	$2.42 \times 10^{-7} \pm 4\,\%$
Path Sampling		f_A	$\mathcal{P}_A(\lambda_n \vert \lambda_0)$	$\mathrm{k} = f_A \times \mathcal{P}_A(\lambda_n \vert \lambda_0)$
TIS		$0.263 \pm 1\%$	$1.52 \cdot 10^{-6} \pm 20\%$	$4.02 \cdot 10^{-7} \pm 20\%$
PPTIS		$0.263 \pm 1\%$	$1.04 \cdot 10^{-6} \pm 19\%$	$2.73 \cdot 10^{-7} \pm 19\%$
RETIS		$0.265 \pm 1\%$*	$1.05 \cdot 10^{-6} \pm 25\%$*	$2.79 \cdot 10^{-7} \pm 25\%$*
FFS (long MD run)		$0.263 \pm 1\%$	$4.69 \cdot 10^{-8} \pm 6\%$*	$1.23 \cdot 10^{-8} \pm 6\%$*
FFS (short MD run)		$0.259 \pm 2\%$	$8.45 \cdot 10^{-9} \pm 9\%$*	$2.18 \cdot 10^{-9} \pm 9\%$*

Final errors were obtained by block averaging and error-propagation rules. The errors of RETIS and FFS are given a star as these errors should not be considered exact due to the neglect of covariant terms that arise due the correlations between path ensembles and initial MD simulation. The FFS was repeated using a shorter (4,000,000 instead of 10,000,000 timesteps) initial MD run.

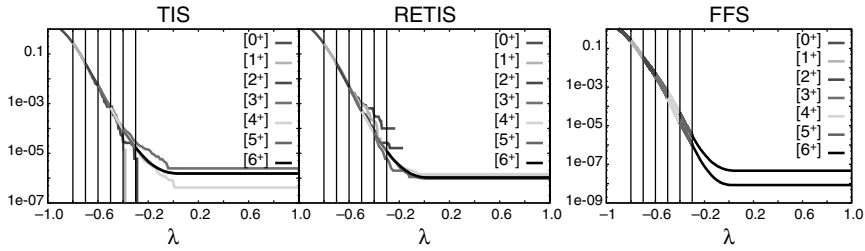


Figure 7. Overall crossing probabilities using TIS, RETIS, and FFS.

compared to Kramer's expression $\kappa \approx (1/\omega_b)(-\gamma/2 + \sqrt{\gamma^2/4 + \omega_b^2}) = 0.9$ with $\omega_b = \sqrt{k_2/m} = \sqrt{2}$. If we compare the TIS, PPTIS, and RETIS simulations we see that they are all close (within a factor 2) to the RF result. The TIS result is somewhat too high that is probably due to a single path ensemble calculation that was not fully converged after 20,000 steps. The RETIS results are clearly better despite the apparent errors that are somewhat smaller for TIS. The RETIS result is much closer to the RF reference. Moreover, it uses only halve the number of shooting moves compared to TIS, which is the most expensive move for realistic systems as it requires a large number of force evaluations. Also, the construction of the overall crossing probabilities in Fig. 7 shows a much better matching between the different ensembles in the RETIS method. The PPTIS result is also very close to the reference value. The PPTIS approximation, Eq. (32), becomes not only exact for very diffusive systems, but is also exact for steeply increasing barriers as all trajectories from λ_i to λ_{i+1} come directly from λ_0 in the past. FFS on the other hand, that is in principal exact unlike PPTIS, gives an unacceptable value that is about a factor 20 too low. Still, if we calculate the error using standard error propagation rules without taking care of the correlations between the ensembles [62], we get errors that seem very low. I also repeated the FFS simulation only changing the length of the initial MD simulation. A decrease of 60% for the initial MD simulation resulted in a final result that is again five times smaller. For TIS, PPTIS, and RETIS the impact of this MD reduction would not even be noticed as it only effects the error in the flux term that is negligible compared to the error in the crossing probability. Figure 7 shows that the two different FFS crossing probabilities are similar at the start, but then start to deviate exponentially. The reason for this behavior is that the true set of reactive trajectories have an average kinetic energy distribution at the start that is considerably shifted compared to the equilibrium distribution. Therefore, a too short MD simulation might not generate sufficient crossing points having a high velocity. As result, the FFS trajectories mainly climb up the barrier helped by the stochastic force instead of a high initial velocity. In Fig. 8, we compare five randomly selected crossing trajectories for TIS

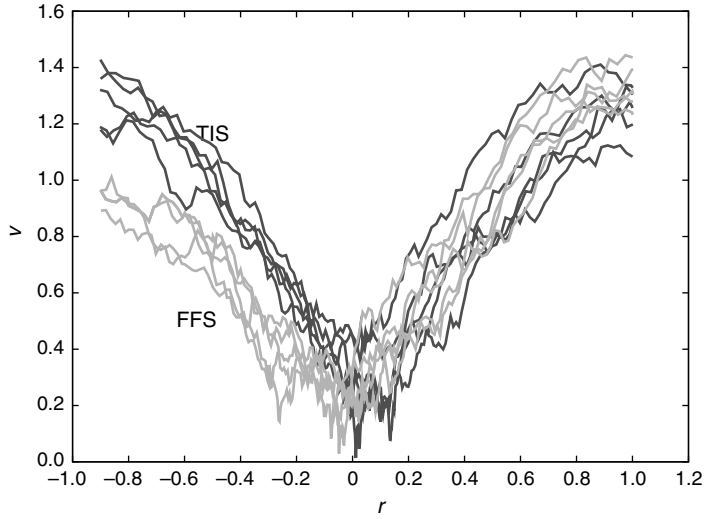


Figure 8. TIS and FFS trajectories in the (r, v) plane.

and for FFS. The FFS trajectories are clearly unrealistic as they are not symmetric in the (r, v) plane, which should be the case for a symmetric barrier. The velocities at the start are much lower than at the other side of the barrier. Moreover, two of the five trajectories, that were randomly picked from the 166 successful ones, start exactly from the same MD crossing point (from a total of 5260 crossing points). TIS, that is able to relax the history of the path, does not show this artifact.

These results have important consequences. It shows that attempts to use FFS for deterministic MD [59, 60] by applying some small level of stochastic noise can only work when inertia effects are not important. In other words, when the deterministic dynamics behaves effectively Brownian. This sampling problem is not unique to FFS, but to any splitting-type method such as weighted ensemble Brownian dynamics [72, 73], Russian Roulette [57, 58], vector walking [74], and S-PRES [75], in which the equations of motion are only followed forward in time. There is not an easy solution for this problem. Still, this is very much desired for nonequilibrium systems for which there are no good alternatives. Possibly, the original nonequilibrium TPS approach of Crooks and Chandler [76] could do better in this situation as it continues rebuilding trajectories from the beginning. Still, the set of starting points follows from a straightforward MD trajectory. Therefore, also this approach is likely to miss these rare initial points that have high potential to become reactive. Time-reversal moves might alleviate the problem [10], but can only be applied for time-reversible dynamics and velocity-symmetric steady-state distributions. The other possibility is to adapt the RC, $\lambda(x)$, to the phasepoint

committor, for instance via an on-the-fly optimization scheme [66]. Such an approach would need to be carried out in full phase space, which is presently not the standard.

However, redefining the intermediate interfaces to lie on the isocommittor surfaces alone will not be sufficient. If we keep the r-dependent definition for stable state A, the committor probability can jump from zero to one is a single timestep (at the point when it leaves state A with a high velocity). Hence, also the state definition A should be redefined in phase space. For this particular system it seems intuitive to use constant energy curves $\lambda(x) = 1/2mv^2 + V(r)$ as RC. FFS will probably work in that case. However, it is yet unclear if it is practically feasible to design appropriate RCs using full phase space in more complex systems. Present algorithms [14] have always assumed that is sufficient to use configuration dependent committor functions.

On the other hand, the TIS methods seems to work properly using configuration dependent RCs. Only to ensure the stability of state A it is sometimes convenient to let λ_0 be velocity dependent [8] (For this system, the friction coefficient is sufficiently high to neglect kineticaly correlated recrossing). On the contrary, it seems that TIS and its variations do not necessarily improve when the RC equals the true isocommittor, a hypothesis that was postulated in Ref. 54. If interfaces are placed at constant energy curves, the trajectories will become much longer than in the present case.

IX. CONCLUSIONS

I have reviewed some dynamical rare event simulation techniques. The RF method is likely the most efficient approach when studying low dimensional systems for which an appropriate RC can easily be found. The most efficient implementation of the RF approach to calculate the dynamical factor is probably the EPF algorithm that is considerably more efficient than the more common transmission coefficient calculation schemes. However, even with this more efficient EPF approach, the RF efficiency will decrease exponentially with barrier height and inverse temperature, if a proper RC cannot be found [39]. The TPS reactive trajectory sampling does not require a RC. However, a definition of a RC is still needed in the TPS rate calculation algorithm, which has been improved by the TIS and RETIS methodologies. The TPS/TIS/RETIS efficiency only scales quadratically with barrier height and inverse temperature when using an "improper" RC [39]. The RC insensitivity of these methods gives them a strong advantage compared to RF methods in complex systems. Of these methods, RETIS is significantly faster than the other two. However, its implementation is somewhat more difficult than TIS. PPTIS (and the similar milestoning) is not an exact method as it assumes memory loss beyond a traveling distance between two interfaces. Using this approximation, PPTIS is able

to reduce the required path length considerably, which can be important for diffusive barrier crossings. FFS does not require information on the phasepoint density and is, therefore, ideally suited to study nonequilibrium events. The advantageous implementation and its apparent efficiency have made FFS a very popular method for equilibrium systems as well. However, the numerical study, presented here, shows that FFS has certain pitfalls that have not yet been reported. It shows that the RC sensitivity of FFS is even more troublesome than it is for RF methods. Present simulation studies have almost always assumed that RC are functions of configuration space alone. My example shows that an appropriate RC for FFS needs to be defined phase space while configurational space would be sufficient for the RF method and the equilibrium path sampling algorithms TPS/(RE)TIS. Still, there are presently no alternative methods that can treat nonequilibrium processes and do not have the same problem. The fact that FFS and other forward MC methods get so easily trapped toward unfavorable reaction paths, by missing an important orthogonal coordinate or velocity in an early stage, requires the uppermost caution when applying these methods and interpreting their results. In addition, this article poses challenges for developing improved nonequilibrium path sampling methods that are either less sensitive to a chosen RC or are able to find an appropriate phase space-dependent RC on the fly.

Acknowledgments

TSvE acknowledges the Flemish Government for long-term structural support via the center of excellence (CECAT) and Methusalem funding (CASAS) and the support of the Interuniversity Attraction Pole (IAP-PAI).

List of Abbreviations

EPF	effective positive flux
FFS	forward flux sampling
MC	Monte Carlo
MD	molecular dynamics
PPTIS	partial path transition interface sampling
RC	reaction coordinate
RE	replica exchange
RETIS	replica exchange transition interface sampling
RF	reactive flux method
TI	thermodynamic integration
TIS	transition interface sampling
TPS	transition path sampling
TS	transition state
TST	transition state theory
US	umbrella sampling

References

1. D. Chandler, *Introduction to Modern Statistical Mechanics*. Oxford University, New York, 1987.
2. D. Frenkel and B. Smit, *Understanding Molecular Simulation*, 2nd ed., Academic Press, San Diego, CA, 2002.
3. C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, Transition path sampling and the calculation of rate constants, *J. Chem. Phys.*, **108**, 1964 (1998).
4. C. Dellago, P. G. Bolhuis, and D. Chandler, Efficient transition path sampling: application to Lennard–Jones cluster rearrangements. *J. Chem. Phys.*, **108**, 9236 (1998).
5. C. Dellago, P. G. Bolhuis, and D. Chandler, On the calculation of reaction rate constants in the transition path ensemble, *J. Chem. Phys.*, **110**, 6617–6625 (1999).
6. P. G. Bolhuis, D. Chandler, C. Dellago, and P.L. Geissler, Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.*, **53**, 291–318 (2002).
7. C. Dellago, P. G. Bolhuis, and P. L. Geissler, Transition path sampling. *Adv. Chem. Phys.*, **123**, 1–78 (2002).
8. T. S. van Erp, D. Moroni, and P. G. Bolhuis, A novel path sampling method for the sampling of rate constants. *J. Chem. Phys.*, **118**, 7762–7774 (2003).
9. T. S. van Erp, Reaction rate calculation by parallel path swapping. *Phys. Rev. Lett.*, **98**(26), 268301 (2007).
10. P. G. Bolhuis, Rare events via multiple reaction channels sampled by path repl ica exchange. *J. Chem. Phys.*, **129**, 114108 (2008).
11. D. Moroni, P. G. Bolhuis, and T. S. van Erp, Rate constants for diffusive processes by partial path sampling. *J. Chem. Phys.*, **120**, 4055–4065 (2004).
12. A. K. Faradjian and R. Elber, Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.*, **120**, 10880 (2004).
13. R. J. Allen, P. B. Warren, and P. R. ten Wolde, Sampling rare switch ing events in biochemical networks. *Phys. Rev. Lett.*, **94**, 018104 (2005).
14. C. Dellago and P. G. Bolhuis, "Transition path sampling and other advanced simulation techniques for rare events," in *Advances in Polymer Sciences*, Vol. 221 on *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, edited by C. Holm and K. Kremer (Springer, Berlin, 2008), pp. 167–233.
15. F. A. Escobedo, E. E. Borrero, and J. C. Araque, Transition path sampling and forward flux sampling applications to biological systems. *J. Phys.-Condes. Matter*, **21**, 333101 (2009).
16. R. J. Allen, C. Valeriani, and P. R. ten Wolde, Forward flux sampling for rare event simulations. *J. Phys.-Conde. Matter*, **21**, 463102 (2009).
17. P. G. Bolhuis and C. Dellago, Trajectory-based rare event simulations, in *Reviews in Computational Chemistry*, San Diego, CA, 2010.
18. P. Hanggi, P. Talkner, and M. Borkovec, Reaction-rate theory—50 years after Kramers. *Rev. Mod. Phys.*, **62**, 251–341 (1990).
19. B. J. Alder and T. E. Wainwright, Phase transition for a hard sphere system. *J. Chem. Phys.*, **27**, 1208 (1957).
20. H. Eyring, The activated complex in chemical reactions. *J. Chem. Phys.*, **3**, 107 (1935).
21. E. Wigner, The transition state method. *Trans. Faraday Soc.*, **34**, 29 (1938).
22. J. C. Keck, Statistical investigation of dissociation cross-section for diatoms. *Discuss. Faraday Soc.*, **33**, 173 (1962).

23. C. H. Bennett, Molecular dynamics and transition state theory: the simulation of infrequent events, in *Algorithms for Chemical Computations*, in R. E. Christofferson, ed., ACS Symposium Series No. 46, American Chemical Society, Washington, D.C., 1977.
24. D. Chandler, Statistical-mechanics of isomerization dynamics in liquids and transition-state approximation. *J. Chem. Phys.*, **68**, 2959–2970 (1978).
25. T. Yamamoto, Quantum statistical mechanical theory of the rate of exchange chemical reactions in the gas phase. *J. Chem. Phys.*, **33**, 281 (1960).
26. J. Horiuti, On the statistical mechanical treatment of the absolute rate of the chemical reaction. *Bull. Chem. Soc. Jpn*, **13**, 210 (1938).
27. G. M. Torrie and J. P. Valleau, Monte-Carlo free-energy estimates using non-Boltzmann sampling—application to subcritical Lennard–Jones fluid. *Chem. Phys. Lett.*, **28**, 578 (1974).
28. E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.*, **156**, 472 (1989).
29. D. Moroni, *Efficient sampling of rare events*, PhD thesis, Universiteit van Amsterdam, 2005.
30. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, The weighted histogram analysis method for free-energy calculations on biomolecules 1. The method. *J. Comput. Chem.*, **13**(8) 1011–1021 (1992).
31. W. K. den Otter and W. J. Briels, The calculation of free-energy differences by constrained molecular-dynamics simulations. *J. Chem. Phys.*, **109**(11), 4139–4146 (1998).
32. H. Grubmüller, Predicting slow structural transitions in macromolecular systems—conformational flooding. *Phys. Rev. E*, **52**, 2893–2906 (1995).
33. E. Darve and A. Pohorille. Calculating free energies using average force. *J. Chem. Phys.*, **115**(20), 9169–9183 (2001).
34. F. G. Wang and D. P. Landau, Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.*, **86**, 2050–2053 (2001).
35. A. Laio and M. Parrinello, Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA*, **99**, 12562 (2002).
36. J. B. Anderson, Predicting rare events in molecular dynamics. *Adv. Chem. Phys.*, **91**, 381–431 (1995).
37. T. S. van Erp, *Solvent effects on chemistry with alcohols*, PhD thesis, Universiteit van Amsterdam, 2003.
38. E. Vanden-Eijnden and F. A. Tal, Transition state theory: variational formulation, dynamical corrections, and error estimates. *J. Chem. Phys.*, **123**, 184103 (2005).
39. T. S. van Erp, Efficiency analysis of reaction rate calculation methods using analytical models I: the two-dimensional sharp barrier. *J. Chem. Phys.*, **125**(17), 174106 (2006).
40. T. S. van Erp and P. G. Bolhuis, Elaborating transition interface sampling methods. *J. Comput. Phys.*, **205**, 157–181 (2005).
41. G. W. N. White, S. Goldman, and C. G. Gray, Test of rate theory transmission coefficients algorithms. an application to ion channels. *Mol. Phys.*, **98**, 1871–1885 (2000).
42. M. J. Ruiz-Montero, D. Frenkel, and J. J. Brey, Efficient schemes to compute diffusive barrier crossing rates. *Mol. Phys.*, **90**, 925–941 (1997).
43. T. S. van Erp and E. J. Meijer, Proton assisted ethylene hydration in aqueous solution. *Angew. Chem.*, **43**, 1659–1662 (2004).
44. K. V. Klenin and W. Wenzel, A method for the calculation of rate constants from stochastic transition paths. *J. Chem. Phys.*, **132**(10), 104104 (2010).

45. J. Rogal and P. G. Bolhuis, Multiple state transition path sampling. *J. Chem. Phys.*, **129**(22), 224107 (2008).

46. D. Ryter, On the eigenfunctions of the Fokker–Planck operator and of its adjoint. *Physica A*, **142**(1-3), 103–121 (1987).

47. B. Peters, G. T. Beckham, and B. L. Trout, Extensions to the likelihood maximization approach for finding reaction coordinates. *J. Chem. Phys.*, **127**, 034109 (2007).

48. E. Vanden-Eijnden, Transition path theory, in M. Ferrario, G. Ciccotti, and K. Binder, eds., *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology*, Volume 703 of *Lecture Notes in Physics*, Springer, New York, US, 2006, pp. 453–493.

49. W. E and E. Vanden-Eijnden, Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events, Volume 61 of *Annual Review of Physical Chemistry*, 2010, pp. 391–420. Publisher = Annual Reviews.

50. A. N. Drozdov and S. C. Tucker, An improved reactive flux method for evaluation of rate constants in dissipative systems. *J. Chem. Phys.*, **115**(21), 9675–9684 (2001).

51. T. S. Van Erp, T. P. Caremans, C. E. A. Kirschhock, and J. A. Martens, Prospects of transition interface sampling simulations for the theoretical study of zeolite synthesis. *Phys. Chem. Chem. Phys.*, **9**(9), 1044–1051 (2007).

52. D. T. Gillespie, Monte-Carlo simulation of random-walks with residence time-dependent transition-probability rates. *J. Comput. Phys.*, **28**, 395 (1978).

53. A. Aerts, M. Haouas, T. P. Caremans, L. R. A. Follens, T. S. van Erp, F. Taulelle, J. Vermant, J. A. Martens, and C. E. A. Kirschhock, Investigation of the mechanism of colloidal silicalite-1 crystallization by using DLS, SAXS, and Si-29 NMR spectroscopy. *Chem.-Eur. J.*, **16**, 2764–2774 (2010).

54. E. Vanden-Eijnden, M. Venturoli, G. Ciccotti, and R. Elber, On the assumptions underlying milestoning. *J. Chem. Phys.*, **129**(17), 174102 (2008).

55. D. Moroni, T. S. van Erp, and P. G. Bolhuis, Simultaneous computation of free energies and kinetics of rare events. *Phys. Rev. E*, **71**, 056709 (2005).

56. P. Majek and R. Elber, Milestoning without a reaction coordinate. *J. Chem. Theory Comput.*, **6**, 1805–1817 (2010).

57. T. E. Booth and J. S. Hendricks, Importance estimation in forward Monte-Carlo calculations. *Nucl. Technol.-Fus.*, **5**(1), 90–100 (1984).

58. P. G. Melnik-Melnikov and E. S. Dekhtyaruk, Rare events probabilities estimation by "Russian Roulette and Splitting" simulation technique. *Probab. Eng. Eng. Mech.*, **15**(2), 125–129 (2000).

59. Z.-J. Wang, C. Valeriani, and D. Frenkel, Homogeneous bubble nucleation driven by local hot spots: a molecular dynamics study. *J. Phys. Chem. B*, **113**(12), 3776–3784 (2009).

60. C. Velez-Vega, E. E. Borrero, and F. A. Escobedob, Kinetics and reaction coordinate for the isomerization of alanine dipeptide by a forward flux sampling protocol. *J. Chem. Phys.*, **130**(22), 225101 (2009).

61. P. G. Bolhuis, Transition path sampling on diffusive barriers. *J. Phys. Cond. Matter*, **15**, S113–S120 (2003).

62. R. J. Allen, D. Frenkel, and P. R. ten Wolde, Forward flux sampling-type schemes for simulating rare events: efficiency analysis. *J. Chem. Phys.*, **124**, 194111 (2006).

63. J. Juraszek and P. G. Bolhuis, Rate constant and reaction coordinate of Trp-Cage folding in explicit water. *Biophys. J.*, **95**(9), 4246–4257 (2008).

64. R. P. Sear, Nucleation in the presence of slow microscopic dynamics. *J. Chem. Phys.*, **128**(21), 214513 (2008).

65. E. Sanz, C. Valeriani, D. Frenkel, and M. Dijkstra, Evidence for out-of-equilibrium crystal nucleation in suspensions of oppositely charged colloids. *Phys. Rev. Lett.*, **99**(5), 055501 (2007).

66. E. E. Borrero, L. M. Contreras Martinez, M. P. DeLisa, and F. A. Escobedo, Kinetics and reaction coordinates of the reassembly of protein fragments via forward flux sampling. *Biophys. J.*, **98**(9), 1911–1920 (2010).

67. T. S. van Erp, Efficient path sampling on multiple reaction channels. *Comput. Phys. Commun.*, **179**, 34–40 (2008).

68. R. H. Swendsen and J. S. Wang, Replica Monte-Carlo simulation of spin-glasses. *Phys. Rev. Lett.*, **57**, 2607–2609 (1986).

69. E. Marinari and G. Parisi, Simulated tempering—a new Monte-Carlo scheme. *Europhys. Lett.*, **19**(6), 451–458 (1992).

70. T. J. H. Vlugt and B. Smit, On the efficient sampling of pathways in the transition path ensemble. *Phys. Chem. Comm.*, **2**, 1–7 (2001).

71. B. Peters and B. L. Trout, Obtaining reaction coordinates by likelihood maximization. *J. Chem. Phys.*, **125**(5), (2006).

72. G. A. Huber and S. Kim, Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.*, **70**(1), 97–110 (1996).

73. B. W. Zhang, D. Jasnow, and D. M. Zuckerman, Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin. *Proc. Natl. Acad. Sci. USA*, **104**(46), 18043–18048 (2007).

74. S. Tănase-Nicola and J. Kurchan, Topological methods for searching barriers and reaction paths. *Phys. Rev. Lett.*, **91**, 188302 (2003).

75. J. T. Berryman and T. Schilling, Sampling rare events in nonequilibrium and nonstationary systems. *J. Chem. Phys.*, **133**, 244101 (2010).

76. G. E. Crooks and D. Chandler, Efficient transition path sampling for nonequilibrium stochastic dynamics. *Phys. Rev. E*, **64**, 026109 (2001).

CONFOCAL DEPOLARIZED DYNAMIC LIGHT SCATTERING

MARCO A.C. POTENZA,[1] TIZIANO SANVITO,[1] VITTORIO DEGIORGIO,[2] AND MARZIO GIGLIO[1]

[1]*Dipartimento di Fisica, Università di Milano, 20133 Milano, Italy*
[2]*Dipartimento di Elettronica, Università di Pavia, 27100 Pavia, Italy*

CONTENTS

I. INTRODUCTION

Particles in solution exhibit both translational and rotational diffusion. The translational diffusion is routinely employed to derive accurate estimates of the size of the scattering particles through the determination of the mass diffusion coefficient D and using the Stokes–Einstein relation $D = k_B T/6\pi\eta R$ [1, 2], the standard technique being dynamic light scattering [3]. Often the experimental procedure involves measurement at different scattering wavevectors $q = (4\pi/\lambda)\sin(\theta/2)$, where θ is the scattering angle and λ the wavelength. The value of D is then derived by determining the diffusive time constants at various wavevectors q, and fitting

Advances in Chemical Physics, Volume 151: Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanoscale Materials, First Edition. Edited by Gregoire Nicolis and Dominique Maes.

them with the relation $\tau_D = 1/(2q^2D)$. The dynamic scattering method for the translational diffusive processes is by now a very well established and widely used method for various applications in physics, chemistry, biology, and medical applications.

Much less known and used is the rotational diffusion. Random kicks from solvent molecules do impose random rotations as well translations to the scattering particle, and while translational diffusion time constants depend on the scattering angle, rotational diffusion has one specific value $\tau_R = 1/(6\Theta) = (8\pi\eta R^3)/(6k_BT)$ and quite remarkably, it coincides with the time constant associated to a translational diffusion over a distance equal to the particle diameter. Rotational diffusion via a scattering experiment can be performed only if the scattering particle has an optical anisotropy, either induced by its form, or by the birefringence of the material of the particle itself. It turns out that the index of refraction mismatch for the anisotropic part is in general vastly smaller than the isotropic part, and therefore the intensity of the depolarized scattered light is very weak compared to the polarized part. Apart from weakness of the signal, there is an additional difficulty due to the inevitable presence of multiple scattering. Indeed, even with perfectly isotropic scattering particles, multiple scattering generates depolarized contributions. Because multiple scattering involves subsequent scattering processes, the associated time constants can be in the range of the expected depolarized components [4], and this is one of the greatest difficulties that prevent the exploitation of depolarized scattering.

One way out has been in the past to adjust the index of refraction of the solvent so it matches the average refractive index of the anisotropic particle, thus drastically reducing the polarized component, and making the procedure viable [5–7]. For water solutions, this limits the method to fluorinated particles, whose average index of refraction is reasonably close to that of water, and addition of urea or similar liquids allows best matching.

In this work, we will present in some details the advantages offered by a scheme that we have recently introduced [8] to drastically reduce the ill effects due to multiple scattering in connection with depolarized dynamic scattering. The method minimizes the region from which the scattered light is fed to the sensor via a tight confocal arrangement. This makes it possible to work under nonindex matching conditions, and the comparison with index matched case is discussed in detail. This possibility opens new perspectives in the study of high concentration anisotropic particles suspended in a fluid, and the application to the study of the first stages of crystal growth is now in progress.

A coverage of basic working formula will be given together with some more recent theoretical results [9]. Possibility of extracting estimates for the translational diffusion from number fluctuations in the very small scattering volume will also be given.

II. FUNDAMENTALS OF DEPOLARIZED SCATTERING: THE STATE OF THE ART

Following the same formalism in Ref. 3, we will refer to the coordinate system shown in Fig. 1.

Coordinates x, y, z identify the laboratory reference system. Axis x is chosen to be along the wavevector of the scattered beam, k_s, while the wavevector k_i of incoming beam is in the plane xy. The polarization plane of the incoming light is assumed to be parallel to z-axis, which will be referred to as the "vertical" (V) direction. The scattered light can be selected through an analyzer to be either parallel or perpendicular to the vertical polarization, thus defining the two scattered fields E_{VV} and E_{VH}, respectively, the so called *polarized* and *depolarized* scattered fields.

The scatterer is assumed to have two components of the polarizability tensor (with respect to the solvent), $\alpha_{||}$ and $\alpha_{\perp}$, respectively, parallel and perpendicular to the symmetry axis. The average polarizability and the anisotropy of the particle polarizability are usually defined as follows:

$$\alpha = \frac{1}{3}\left(\alpha_{||} + 2\alpha_{\perp}\right) \tag{1}$$

$$\beta = \left(\alpha_{||} - \alpha_{\perp}\right) \tag{2}$$

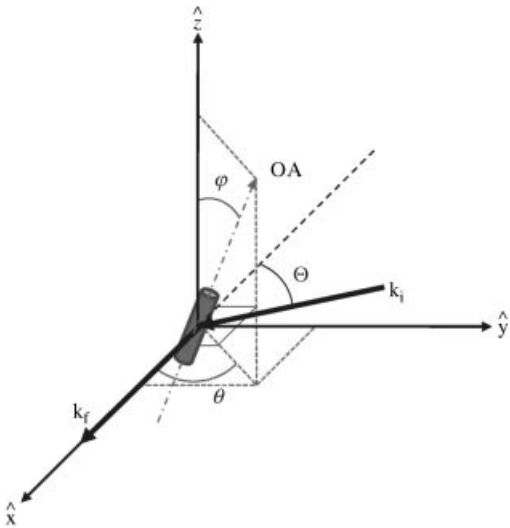


Figure 1. The coordinate system adopted in the text: XY is the scattering plane, Θ is the scattering angle, angles θ, φ identify the orientation of the optical axis (OA).

For a particle with optical axis oriented along the direction θ, φ as depicted in Fig. 1, the resulting components of the polarizability tensor are then given by [3]

$$\alpha_{zz} = \alpha + \sqrt{\frac{16\pi}{45}}\beta Y_{2,0}(\theta, \varphi) \tag{3}$$

$$\alpha_{yz} = j\sqrt{\frac{2\pi}{15}}\beta\left[Y_{2,1}(\theta, \phi) + Y_{2,-1}(\theta, \varphi)\right] \tag{4}$$

As shown in Ref. 3, if a sample containing N scatterers is illuminated by a plane wave polarized along the z, vertical axis, and the scattered light emitted at a given angle θ is measured either with polarization along axis z (VV) or with polarization along axis y (VH), then the following temporal correlation functions can be obtained:

$$\begin{aligned} G_{\mathrm{VV}}(\mathbf{q}, t) &= \langle N\rangle\alpha^2 e^{-q^2 Dt} + \frac{4}{45}\langle N\rangle\beta^2 e^{-6\Theta t}e^{-q^2 Dt} \\ &= G_{\mathrm{ISO}}(\mathbf{q}, t) + \frac{4}{3}G_{\mathrm{VH}}(\mathbf{q}, t) \end{aligned} \tag{5}$$

$$G_{\mathrm{VH}}(\mathbf{q}, t) = \frac{1}{15}\langle N\rangle\beta^2 e^{-6\Theta t} e^{-q^2 Dt} \tag{6}$$

where, as mentioned above, VV and VH indicate the mutual polarization directions of the incoming beam and the measured scattered light, $\mathbf{q} = \mathbf{k_i} - \mathbf{k_f}$ is the transferred momentum, D is the mass diffusion coefficient:

$$D = \frac{k_{\mathrm{B}}T}{6\pi\eta R} \tag{7}$$

and Θ is the rotational diffusion coefficient:

$$\Theta = \frac{k_{\mathrm{B}}T}{8\pi\eta R^3} \tag{8}$$

It is worth noticing that the first term in Eq. (3) solely depends on the isotropic part of the polarizability tensor, thus being independent of the scatterer orientation. This term is due only to *mass diffusion*, and it is the only contribution surviving in the simpler case of optically isotropic scatterers.

Both Eqs. (5) and (6) contain terms depending on the temporal changes of the optical axis orientation, decaying with a time constant given by Θ. This terms describes the *rotational diffusion*. Apart from inessential numerical factors, the amplitude of this terms compared to the polarized term in Eq. (5) scales as β^2/α^2 that is typically smaller that 10^{-4}.

For short correlation times, Eqs. (5) and (6) obtained above also provide two important formulas relating just the coefficient α and β:

$$G_{\mathrm{ISO}} = \langle N \rangle \alpha^2 \tag{9}$$

$$G_{\mathrm{VH}} = \frac{1}{15} \langle N \rangle \beta^2 \tag{10}$$

Finally, it can be easily proved that

$$G_{\mathrm{VV}} = G_{\mathrm{ISO}} + \frac{4}{3} G_{\mathrm{VH}} \tag{11}$$

and

$$\frac{G_{\mathrm{VH}}}{G_{\mathrm{ISO}}} = \frac{1}{15} \frac{\beta^2}{\alpha^2} = \frac{3}{5} \left[\frac{\alpha_{||} - \alpha_{\perp}}{\alpha_{||} + 2\alpha_{\perp}} \right]^2 \tag{12}$$

meaning that the ratio β^2/α^2 can be measured from both the correlation functions of the isotropic diffusion [see Eq. (5)] and for the VH component [Eq. (6)].

Especially for small particles, the main contribution in Eq. (6) is given by the depolarized component. That's why it becomes imperative to collect the I_{VH} scattered light to measure rotational diffusion coefficients.

III. THE TRADITIONAL APPROACH

A peculiar geometry used in the past [7], takes advantage from heterodyning the zero-angle scattered field by using the beam transmitted by the sample.

The great advantage in exploiting this configuration is related to the time dependence of both the translational and rotational diffusion. As it appears from Eqs. (5) and (6), the translational characteristic times are q dependent, becoming shorter for larger scattering angles, or q. By contrast, the rotational characteristic times are q independent. Therefore, by looking toward the scattering sample at very low angles one will see the scattered field affected by fast, minute fluctuations changing with rotational times, and a much slower modulation of the intensity that eventually will destroy the correlation. Moreover, due to the $1/R^3$ dependence of the rotational diffusion coefficient [Eq. (8)] to be compared to the $1/R$ dependence of the Stokes–Einstein relation [Eq. (7)], by increasing r the rotational characteristic times can easily be comparable to the translational, and thus undistinguishable from them. So one is forced to extend as much as possible the translational correlation times by working at low angle, ideally at zero angle.

Heterodyning the signal by collecting the scattered light at zero angle together with the transmitted beam brings to the following superposition of fields:

$$E = E_{\mathrm{t}} + E_{\mathrm{VV}} + E_{\mathrm{VH}} \tag{13}$$

where E_t is the transmitted beam acting as the local oscillator, and E_{VV}, E_{VH} the two field with polarizations parallel and perpendicular to the polarization plane of the incoming beam (see above).

More precisely, since the scattered field is in quadrature with the transmitted [10], the superposition of fields expressed by Eq. (13) is not effective in changing the intensity. A quarter wave plate inserted downstream the sample was used to the aim, as described in detail in Ref. 7. In such a way the depolarized scattered field is phase shifted by $\pi/2$ and the superposition [Eq. (13)] operates as a destructive interference. The corresponding intensity can be written under the hypothesis that the amplitude of the scattered radiation is much smaller than the transmitted, so that the second-order terms can be neglected. In such a way

$$I(\mathbf{q}=0) = I_t + 2\sqrt{I_t}\Re[E_{VV}] + 2\sqrt{I_t}\Re[E_{VH}] \tag{14}$$

where $\Re$ indicates the real part of the complex numbers.

Notice that the polarization plane of the field E_{VV} is the same of the much more intense E_t. This prevents for any measure of this depolarized component. The only depolarized component that can be measured is the E_{VH} that can be selected by crossing an analyzer with the polarization plane of the incoming beam. By accurately changing the axis of the analyzer one can change the amount of transmitted light by a factor f accordingly, thus determining a new local oscillator that has a much smaller amplitude. By contrast, the depolarized intensity is completely transmitted.

The light intensity after the analyzer is given by

$$I'(\mathbf{q}=0) \approx I_{LO} + 2\sqrt{I_{LO}}\Re\left[fE_{VV}\right] + 2\sqrt{I_{LO}}\Re[E_{VH}] \tag{15}$$

where the new local oscillator is $I_{LO} = fI_t$.

IV. THE NOVEL APPROACH

The method exploits recent results obtained about the fundamentals of the optical theorem (OT) [9]. By using the adimensional scattering amplitudes as in Ref. 10, for a collection of N monomers with both positions and orientations at random, the polarized and the depolarized scattered amplitudes are given by

$$S_{VV}(0,t) = \left[ik^3\alpha + \frac{2}{3}k^6\alpha^2 + \frac{4}{27}k^6\beta^2\right]N(t) + \left[ik^3\beta + \frac{2}{9}k^6\left(\alpha_3^2 - \alpha_1^2\right)\right]\sum_{j=1}^{N}\left[\sqrt{\frac{16\pi}{45}}Y_{2,0}(\theta_j(t),\varphi_j(t))\right] \tag{16}$$

$$S_{\mathrm{VH}}(0,t) = \left[ik^3\beta + \frac{1}{3}k^6\left(\alpha_3^2 - \alpha_1^2\right)\right]\sum_{j=1}^{N} \times \left[\sqrt{\frac{2\pi}{15}}\left(Y_{2,1}(\theta_j(t),\varphi_j(t)) + Y_{2,-1}(\theta_j(t),\varphi_j(t))\right)\right] \quad (17)$$

where $k = 2\pi/\lambda$ and λ is the wavelength, φ_j and θ_j are the Euler angles giving the orientation of each monomer j (for fixed time; see Fig. 1) and $Y_{h,k}$ indicate the spherical harmonics functions (see above, and for further details [3]). As it can be immediately appreciated, as monomers change their orientations just the result of the sum changes, thus making the first- and second-order terms fluctuating synchronously. Indeed, one can equally measure rotational diffusion from one of the two terms at will. The first-order term has been considered only for the previous DDLS experiments. It was the only known at the time, and under index matching conditions the first term is by far the leading one, the other being actually negligible. That's imposed to phase shift the depolarized scattered field by $\pi/2$ for making it interfering with the transmitted beam in heterodyne measurements as mentioned above [7]. Here we propose a method exploiting the measurement of the second-order term. Indeed, Eq. (17) shows that the same dynamic of the scattered field could in principle be obtained by heterodyning the second-order scattered component with the transmitted field, with no need for any phase shifting. Incidentally the experimental apparatus becomes simpler than the one needed for studying the index matched samples. Moreover, well out of index matching conditions, the second term increases and, depending on the particle size, can also be predominant. This is due to the peculiar combination of α_1 and α_3, $\alpha_3^2 - \alpha_1^2$, that is not immediately related to β. Just to have an example, a plot of $k^3\beta$ and $k^6\ (\alpha_3^2 - \alpha_1^2)$ is presented in Fig. 2 as a function of the solvent refractive index, for a suspension of colloids 160 nm in diameter. Typical values for the ordinary and extraordinary indexes have been assumed, namely $n_1 = 1.59$ and $n_3 = 1.58$. While the first-order ($k^3\beta$) term is almost constant across the whole index range, the second-order one rapidly increase out of index matching.

Well out of index matching conditions the term S_{VV} can be enormous compared to the S_{VH}, so that the multiply scattered field can be comparable or larger than the depolarized component. The spurious depolarized field generated by multiple scattering can fluctuate with characteristic times fast enough to hide the genuine, fainter depolarized signal [4].

The depolarized light will be heavily dominated by the multiply scattered contributions, thus hiding the much fainter depolarized scattered light. To get rid of it, we have proposed to illuminate the sample with a beam that is strongly focalized, and to collect light from the focal region with a tight confocal scheme. The scattering region is then actually reduced to the superposition of the illuminated

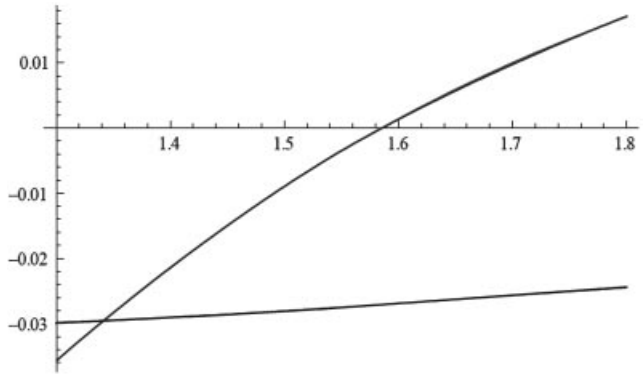


Figure 2. The behavior of $k^3\beta$ and k^6 $(\alpha_3^2 - \alpha_1^2)$ for 160 nm diameter colloids suspended in a solvent with refractive index changing from 1.3 to 1.8. The colloids have been assumed to be endowed with refractive indexes $n_1 = 1.59$ and $n_3 = 1.58$. The first order term is almost constant, across the whole index range, while the second order one significantly increases out of index matching.

focal spot and the region from which light is drawn into the collection fiber. If the light emitted within the scattering region has a mean free path that is much larger than the linear size of the scattering volume, the light propagating along the transmitted beam is just the pure depolarized component to be analyzed (see Fig. 3 for a carton representation of the scattering region). Double or in general multiple scattering events just occur out of the scattering region, and are then effectively gotten rid of. Very concentrated samples can then be used and the second term in Eq. (17) gives a deceivingly measurable contribution to be measured with a traditional photomultiplier for DLS.

By introducing the standard definition for the intensity autocorrelation function:

$$G_2(\tau) = \langle I(0)I(\tau)\rangle \tag{18}$$

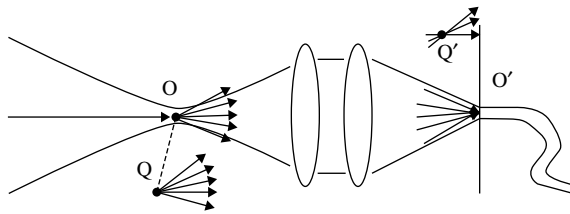


Figure 3. Schematic representation of two typical scattering events. The light emerging from the scattering volume (shaded region) is drawn into the optical fiber. The light originated outside the scattering region is focused out of the fiber, contributing with negligible power to the collected light.

the polarized and depolarized scattered amplitudes [see Eqs. (16) and (17)] can be written as

$$\Re\left[S_{\mathrm{VV}}(t)\right] = \frac{2}{3}k^6\alpha^2 N(t) \tag{19}$$

$$\Re\left[S_{\mathrm{VH}}(t)\right] = \frac{1}{3}k^6\left(\alpha_3^2 - \alpha_1^2\right)\sum_{j=1}^{N}\left[\sqrt{\frac{2\pi}{15}}\left(Y_{2,1}(t) + Y_{2,-1}(t)\right)\right] \tag{20}$$

where next to the leading order contributions are neglected.

The averages and mutual correlations of the polarized and depolarized scattering fields are

$$\langle S_{\mathrm{VH}}(0)\rangle = \langle S_{\mathrm{VH}}(t)\rangle = 0 \tag{21}$$

$$\langle S_{\mathrm{VV}}(0)\rangle = \langle S_{\mathrm{VV}}(t)\rangle = \frac{2}{3}k^6\alpha^2\langle N\rangle \tag{22}$$

$$\langle S_{\mathrm{VV}}(0)S_{\mathrm{VH}}(\tau)\rangle = \langle S_{\mathrm{VV}}(\tau)S_{\mathrm{VH}}(0)\rangle = 0 \tag{23}$$

$$\langle S_{\mathrm{VV}}(0)S_{\mathrm{VV}}(\tau)\rangle = \frac{4}{9}k^{12}\alpha^4\langle N\rangle^2 + \frac{4}{9}k^{12}\alpha^4\langle\delta N(0)\delta N(\tau)\rangle \tag{24}$$

$$\langle S_{\mathrm{VH}}(0)S_{\mathrm{VH}}(\tau)\rangle = \frac{1}{135}k^{12}\left(\alpha_3^2 - \alpha_1^2\right)^2\langle N\rangle e^{-6\Theta\tau} \tag{25}$$

The autocorrelation intensity function is then

$$\begin{aligned} G_2(\tau) = &\left[|E_{\mathrm{t}}|^4 f^2 + \frac{8}{3}|E_{\mathrm{t}}|^3 f^2 k^6\alpha^2\langle N\rangle + \frac{16}{9}|E_{\mathrm{t}}|^2 f^2 k^{12}\alpha^4\langle N\rangle^2\right] \\ &+\frac{16}{9}|E_{\mathrm{t}}|^2 f^2 k^{12}\alpha^4\langle\delta N(0)\delta N(\tau)\rangle + \frac{4}{135}|E_{\mathrm{t}}|^2 f k^{12}\left(\alpha_3^2 - \alpha_1^2\right)^2\langle N\rangle e^{-6\Theta\tau} \end{aligned} \tag{26}$$

In Eq. (26) three terms are present: the first one is a constant, the second brings information about the number fluctuations within the scattering volume, and the third just depends by the rotational diffusion dynamics.

Equation (26) can be normalized for the squared average intensity:

$$\begin{aligned} g_2(\tau) \approx 1 + g_{\mathrm{NF}}(\tau) + g_{\mathrm{ROT}}(\tau) \approx 1 + B\left(\frac{16}{9}f^2\alpha^4\langle\delta N(0)\delta N(\tau)\rangle\right. \\ \left. + \frac{4}{135}f\left(\alpha_3^2 - \alpha_1^2\right)^2\langle N\rangle e^{-6\Theta\tau}\right) \end{aligned} \tag{27}$$

where B is a instrumental constant. Also Eq. (12) becomes

$$\frac{g_{\mathrm{ROT}}(\tau = 0)}{g_{\mathrm{NF}}(\tau = 0)} = \frac{1}{15}\frac{\beta^2}{f\alpha^2} \tag{28}$$

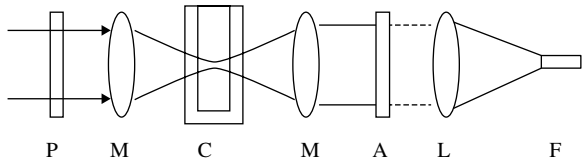


Figure 4. Schematic of the optical layout. P: polarizer; M: microscope objectives; C: cell; A: analyzer; L: focusing lens; F: optical fiber.

which shows that the correlation function also contains the information about the optical anisotropy of the particles.

V. THE OPTICAL LAYOUT

The optical layout is schematically shown in Fig. 4. The beam of a He–Ne laser (He–Ne JDS Uniphase 1100, $\lambda = 632.8$ nm, 20 mW) is sent through a Glan-Thomson polarizer, spatially filtered, expanded and collimated through a lens doublet. Two identical long distance microscope objectives (M: Nachet 20x, NA = 0.30) are used to focus the laser light into the sample cell (C) and for collecting in the forward direction both the transmitted and the scattered light coming from a small region within the sample. A focal spot 1.5 μm in diameter determines the scattering volume, approximately 7.5 μm long. The cell is 2 mm thick with plane parallel walls 2.5 mm thick. The emerging collimated beam passes through an analyzer (A) in conditions of not complete extinction and launched by an aspheric lens (*L*: focal length $f = 45$ mm, NA = 0.55) into a monomode optical fiber (FS-SN-3224, core diameter 4.3 μm) coupled with the cathode of a photomultiplier (model PMT120-OP). The electronic signal is sent to a digital multi-tau correlator (FLEX-02-01-D, Correlator.com, CA), with a minimum lag time of 3.3 ns. It provides a normalized correlation function evaluated in 1120 lag times and a history of the intensity sampled each 0.17 s. The leakage of the main beam through the analyzer acts as a local oscillator, tunable in amplitude by varying the position of the analyzer, heterodyning the weak depolarized scattered light as described in Section II.

VI. DATA REDUCTION SCHEME

A typical correlation curve plotted in a log–log plane is shown in Fig. 5, obtained with a water suspension of depolarized monomers 100 nm in radius at a concentration as described below. Two different modes are evident. The fast mode is determined by the decorrelation in the depolarized light induced by the rotations of the monomers, and provides the rotational characteristic times. To extract the

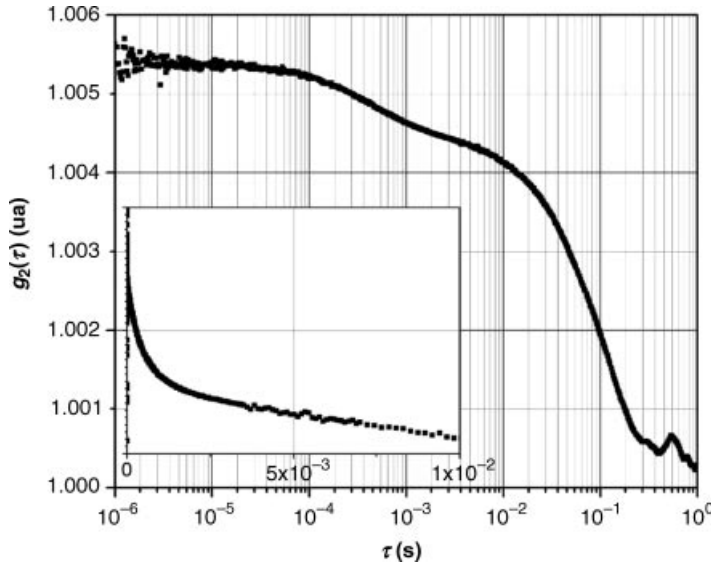


Figure 5. A typical correlation curve obtained with monomers 80 nm in radius.

rotational correlation function we need to separate the two processes, taking care of the slow mode for first and subtracting it to the overall curve.

The slow mode is determined by the fluctuations of the local oscillator, and can be either due to an intensity instability of the laser or to the intensity changes due to the number fluctuations within the scattering volume. The optical system has then been designed to minimize the characteristic times of the number fluctuations. In literature, the number fluctuations traditionally introduced for the simple case of a flat-top illuminating laser beam. Here we present the behavior of the correlation function obtained for a Gaussian beam profile, that is exactly what we have at the scattering volume. It can be easily shown that for small lag times the correlation function for the VV contribution is proportional to the number of monomers within the scattering volume, N, and decays with an exponential behavior:

$$\langle N(0)N(\tau)\rangle \xrightarrow{\tau\to 0} N\, e^{-\tau/\tau_0} \tag{29}$$

where the characteristic time is related to the mass diffusion coefficient D as follows:

$$\tau_0 \approx \frac{{w_0}^2}{2D} \tag{30}$$

The asymptotic behavior in Eq. (29) is more than enough for fitting the slow mode. For longer times the dependence is not straightforward, but can be easily obtained through numerical computation.

The dependence upon N is explained as follows. The power of the extinguished light is proportional to N, and its fluctuation to $N^{1/2}$ because of Poisson's statistics arguments. But the light being collected forward, all the emitted fields sum up coherently, the intensity turning out to be proportional to N.

Notice that the function describing Eq. (26) also contains the leakage fraction f, that determines the relative amplitudes of the two modes in Eq. (28).

The VH contribution is proportional to N as well. The physical origin of this dependence is completely different, being the sum of N random phased scattered fields, each one depending on the corresponding monomer orientation. It results that the relative amplitude of the two modes provides information about the ratio R between $\langle S_{\mathrm{VV}}^2 \rangle - \langle S_{\mathrm{VV}} \rangle^2$ and $\langle S_{\mathrm{VH}}^2 \rangle$, $R = bf\alpha^2/\beta^2$, that is related to the particle anisotropy.

In order to extract the fast mode from the correlation function in Fig. 6, we first need to characterize the slow mode. From the characteristic time of this mode a rough estimate of the mass diffusion coefficient can be attempted.

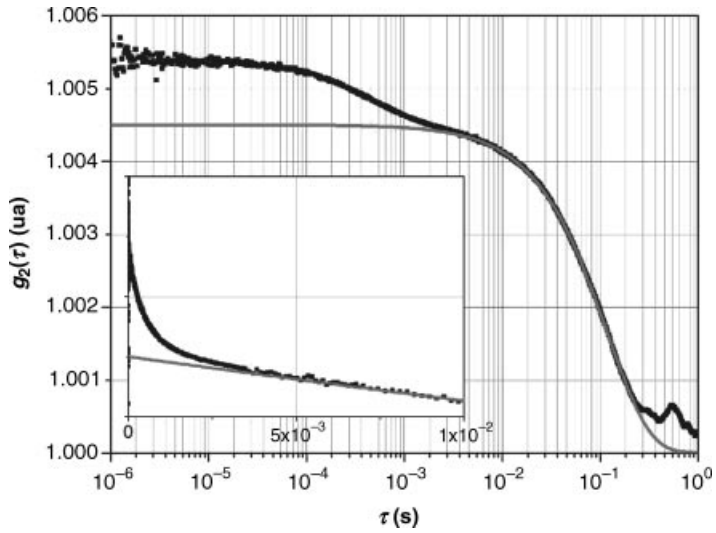


Figure 6. The time correlation function shown in Fig. 5 plotted on a log–log scale. The continuous line shows a best fitting with a single exponential. The insert shows the same plot on a log-lin scale, making evident the exponential decay of the slow mode, providing the characteristic time of the diffusive brownian motions.

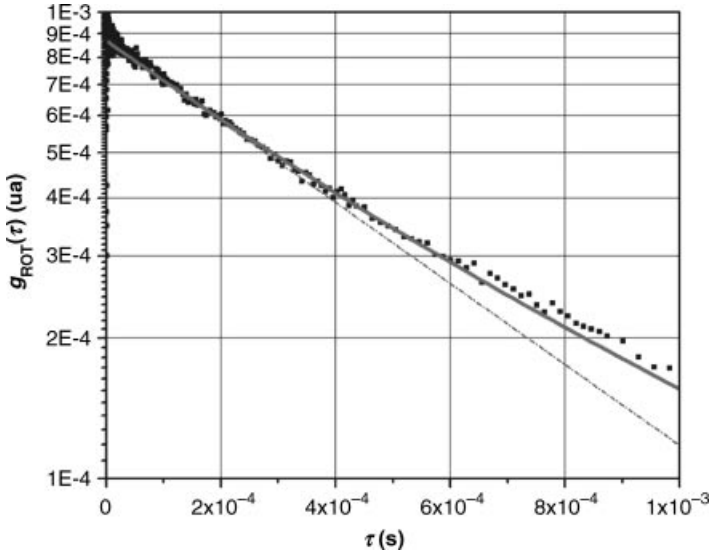


Figure 7. The fast mode obtained from the correlation function subtracted for the slow mode shown in Fig. 6. Notice the strong nonexponential decay due to a small polydispersity of the sample (see text for discussion).

The fast mode is then obtained by subtracting the slow mode from the correlation function, turning out with the curve plotted in Fig. 7. Due to the sample polydispersity it is not a single exponential decay. A rough estimate of the characteristic time can be extracted from the slope of the curve for small lag times. A better description of the curve can be obtained by taking into account the polydispersity through a revisitation of the cumulant theory, applied to the rotational case. A fundamental difference arise from the $1/R^3$ dependence of the rotational diffusion, that introduces a strong asymmetry in the contributions to the correlation curve also with a small, symmetric polydispersity.

Let $W(\Gamma)$ be the distribution of the characteristic times, that is related to the distribution of radii of the monomers, $n(R)$. Then the field correlation function can be described as follows:

$$g_{\mathrm{ROT}}(\tau) = \int_0^\infty W(\Gamma) e^{-\Gamma\tau}\, d\Gamma \tag{31}$$

For a monomodal size distribution, $W(\Gamma)$ can be expanded, and this brings to the following form for the correlation function:

$$g_{\mathrm{ROT}}(\tau) = e^{-6\langle\Gamma\rangle\tau}\left(1 + \frac{\mu_2}{2!}\tau^2 + \cdots\right) \tag{32}$$

where $\langle\Gamma\rangle$ is the characteristic rotational diffusion time Θ, and $\mu_2 = \sigma_\Gamma^2$ is the second cumulant. Keeping into account the dependence of the rotational diffusion coefficient, as shown in Eq. (8), the second cumulant can be related to the radii variance through the following relation:

$$\mu_2 = \sigma_\Gamma^2 \approx \left(\frac{6k_B T}{8\pi\eta}\right)^2 \frac{9}{\langle R\rangle^8}\sigma_r^2. \tag{33}$$

where $\langle R\rangle$ is the average hydrodynamic radius R of the monomers.

VII. RESULTS

To test the reliability of the method we have performed a number of measures with known samples in several experimental conditions. The samples have been chose to be quasimonodisperse water suspensions of spherical particles endowed with strong optical anisotropy. The monomers have been produced by the group of M. Ballouff (see, for example [11–15] and references therein). A PMMA sphere is included in a protein network constituted by PNIPAM, encapsulating much smaller palladium nanospheres giving the optical anisotropy. Due to the thermal properties of the PNIPAM, the volume of the whole network can be changed by changing the temperature T of the system, more precisely it shrinks by increasing temperature. At $T = 32^\circ$C a phase volume occurs. The samples are suspended in a water solution of KCl 0.05 M.

The two samples provided by M. Ballouff have been previously characterized through the Cryo-TEM analysis at many temperatures, and they have been determined to have radii of approximately 80 and 100 nm, with a polydispersity of 9.5% both. Due to the intrinsic high resolution of the rotational measurements, as well as to the temperature dependence due to both the PNIPAM network behavior and the water viscosity changes, accurate temperature measurements have been necessary. Moreover, traditional dynamic light scattering measurements have been performed at 90° by strongly diluting the samples with proper solutions, down to 1.3×10^{-5} w/w and 5.4×10^{-6} w/w, respectively. Data are reported in Fig. 8 (circles and squares) together with exponential best fits that gives the average radii, 78.2 ± 0.9 nm (at 23°C) and 108.4 ± 0.8 nm (at 24°C).

For each measurement a simple turbidity estimate has also been done. The sample cell has been transilluminated by a collimated laser beam and the intensity measured in the far field through a calibrated photovoltaic cell. Attenuation can be varied from a few percent up to 99% under controlled conditions. The data presented below have been obtained with transmissions of approximately 50%. Moreover, an example of the reliability of the method at different concentrations is shown at the end of this section.

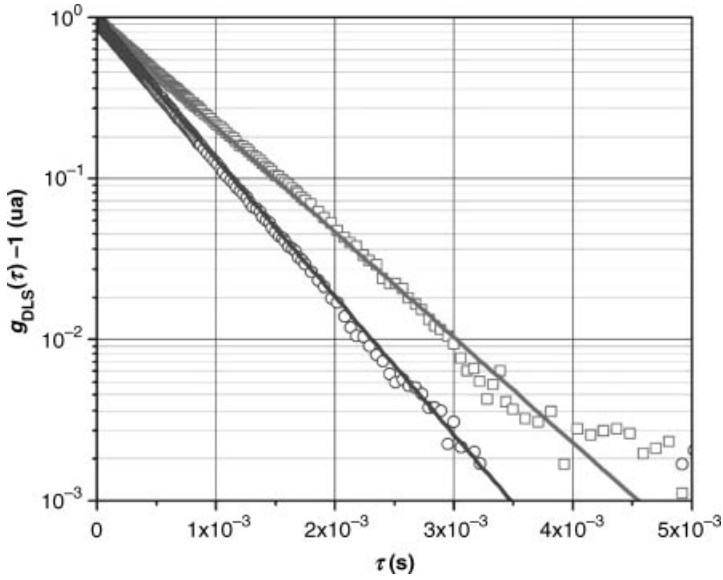


Figure 8. Dynamic Light Scattering measurements of the two samples used to measure the rotational dynamics. Strong dilutions have been used (see text).

The data reduction scheme described in Section V has been used to recover the rotational correlation functions by subtraction of the slow modes. Data are represented in Fig. 9 for the two samples. Experimental points show a relatively small noise upon more than a decade on the vertical axis.

For large lag times the tail of the curve is evidently not exponential, accordingly to the polydisperse nature of the sample. Solid lines are best fit obtained with the cumulants method as previously described. The best fits, corresponding to a polydispersity of approximately 10%, has to be compared to the Cryo-TEM data reported above and to our own traditional DLS results (see below).

The measured hydrodynamic radii resulted to be respectively 79.6 ± 0.6 nm ($T = 23.2^\circ$C) and 106 ± 3 nm ($T = 24.5^\circ$C), to be compared to the Cryo-TEM data at the same temperatures, 79.9 and 102.7 nm, and to the DLS measurements, 78.2 ± 0.9 and 108.4 ± 0.8 nm.

Uncertainties in radii are limited to a few percent, due to the strong dependence of the characteristic times on the radii. Our data seem consistent also with the presence of a non-negligible skewness, but no data dealing with this feature exist to compare our result.

In Table I, we report the results of a series of measurements at different concentrations obtained with the smaller monomers.

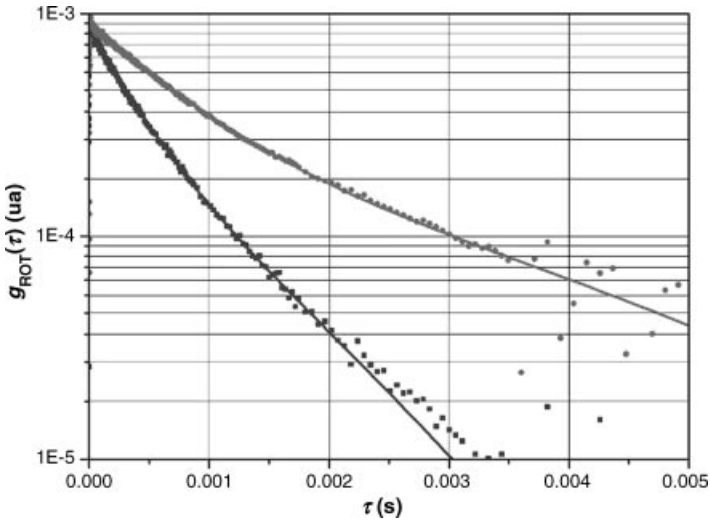


Figure 9. Typical rotational modes obtained with the described method from measurements of known anisotropic monomers. Dots represent experimental points. Solid lines are best fits obtained with the cumulant method (see Section IV).

TABLE I
The Measured Rotational Radii as a Function of the Colloid Volume Fraction

c (Volume Fraction)	R (nm)
1.3×10^{-4}	82 ± 2
2.6×10^{-3}	79 ± 2
6.5×10^{-3}	80 ± 3
9.8×10^{-3}	79 ± 2
1.3×10^{-2}	77 ± 2

The stability within a few percent is found for concentrations ranging upon two decades. Notice that at higher concentrations the contributions from multiple scattering begin to affect the data, by reducing the correlation times. At the lowest concentrations the uncertainties are mainly due to the relative small signals due to the small number of monomers within the small scattering volume. Indeed, this method takes advantage just from the high concentration of the suspension that increases the rotational signal with respect to the power of the local oscillator and then brings to faster measurements.

VIII. CONCLUSIONS

In conclusion we have proved that huge advantages can be obtained by using a heterodyne DDLS method operated in confocal configuration to measure the rotational time constants of highly concentrated colloidal suspensions. This overcomes the typical limitations affecting the traditional DDLS methods. Operated with known samples, the method brings the expected results for the size as well as the polydispersity, determined through the cumulant methods modified to the aim. To the best of our knowledge this possibility has never been possible before, and opens a completely new approach in the study of colloidal suspensions. Besides the precise characterization of radii and crystallinity degree of particles in concentrated suspensions, the method could be applied to the modern approaches to the physics of nonequilibrium phenomena, phase transitions, as well as we foresee the application of this technique operated with known spheres for the study of local viscosity in fluids, especially close to transitions and critical points. Work is in progress to apply the method for studying the first stages of protein crystallization, starting from submicron sized structures that are expected to exhibit a degree of order that can be gauged through relatively simple control mechanisms related to the specific phase diagram of the systems (see, for example [16] and references therein).

Acknowledgment

We are particularly grateful to M. Ballouff for providing the anisotropic samples and the relative information needed to operate with them.

References

1. A. Einstein, *Ann. Phys.*, **17**, 549 (1905); A. Einstein, *Investigations on The theory of the Brownian Movement*, Dover Publications, Inc., 1956.
2. G. G. Stokes, *Cambridge Philos. Soc.*, **9**, 8 (1851).
3. R. Pecora and B. J. Berne, *Dynamic Light Scattering*, Dover Publications, Inc., 1976.
4. R. C. Mockler, C. M. Sorensen, and W. J. O'Sullivan, *Phys. Rev. A*, **14**, 1520 (1976).
5. V. Degiorgio, R. Piazza, T. Bellini, and M. Visca, *Adv. Colloid Interface Sci.*, **48**, 61 (1994).
6. V. Degiorgio, R. Piazza, and R. B. Jones, *Phys. Rev. E*, **52**, 2707 (1995).
7. R. Piazza, F. Mantegazza, V. Degiorgio, and T. Bellini, *Physica A*, **235**, 279 (1997).
8. M. A. C. Potenza, T. Sanvito, M. D. Alaimo, V. Degiorgio, and M. Giglio, *Eur. Phys. J. E*, **31**, 69 (2010).
9. V. Degiorgio, M. A. C. Potenza, and M. Giglio, *Eur. Phys. J. E*, **29**, 379 (2009).
10. H. C. Van de Hulst, *Light Scattering by Small Particles*, Dover Publications Inc., 1957.
11. M. Ballauff, *Dynamics of the Thermosensitive Spherical Microgel and the Validity of Stokes–Einstein Relations*, private communication, 2007.
12. J. J. Crassous, J. Schmidt, M. Ballauff, M. Drechsler, and Y. Talmon, *Langmuir–Am. Chem. Soc.*, **22**, 2403 (2006).
13. J. J. Crassous, M. Siebenbürger, M. Ballauff, M. Drechsler, O. Henrich, and M. Fuchs, *J. Chem. Phys.*, **125**, 204906 (2006).

14. J. J. Crassous, A. Wittemann, M. Siebenbürger, M. Schrinner, M. Drechsler, and M. Ballauff, *Colloid Polym. Sci.*, **286**, 805 (2008).
15. N. Dingenouts, C. Norhausen, M. Ballauff, *Macromolecules* **31**, 8912 (1998).
16. P. Vekilov, *Crystal Growth Design* **4**, 671 (2004).

THE TWO-STEP MECHANISM AND THE SOLUTION–CRYSTAL SPINODAL FOR NUCLEATION OF CRYSTALS IN SOLUTION

PETER G. VEKILOV

Department of Chemical and Biomolecular Engineering and Department of Chemistry, University of Houston, Houston, TX 77204-4004, USA

CONTENTS

Advances in Chemical Physics, Volume 151: Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanoscale Materials, First Edition. Edited by Gregoire Nicolis and Dominique Maes.

I. INTRODUCTION

The nucleation of crystals determines many properties of the emerging crystal population. Since nucleation selects the polymorphic form, if a different polymorph is desired, conditions at which its nucleation is faster than that of the other possible polymorphs should be sought. If nucleation is fast, many crystals form nearly simultaneously. Their growth depletes the solution of solute and may lead to cessation of nucleation at the later stages of crystallization. Thus, the majority of crystals grow to approximately identical sizes. In contrast, if nucleation is slow and fewer crystals nucleate at a time, the supersaturation in the solution drops slowly, the nucleation of new crystals continues and a population of crystals of various sizes forms. Ultimately, if nucleation is hindered everywhere in the growth container but at a few selected spots, crystals only nucleate at these spots and grow large before the solution is depleted of nutrient. Hence, control of nucleation is a means to control size, size distribution, polymorphism, and other properties of the crystals.

Here, we review recent advances in the understanding of nucleation of crystals from solution. Solution crystallization underlies a broad range of industrial, laboratory, and physiological processes. Single solution-grown crystals of inorganic salts or mixed organic–inorganic materials are used in nonlinear optics elements [1] and for other electronic and optical-electronic applications; chemical products, and production intermediates are precipitated as crystals in thousands-of-tons amounts. Another area that relies on solution-grown crystals is pharmacy: the slow crystal dissolution rate is used to achieve sustained release of medications: small-molecules organic [2], or protein such as insulin, interferon-α, or the human growth hormone [2–6]. If the administered dose consists of a few equidimensional crystallites, steady medication release rates can be maintained for longer periods than for doses comprised of many smaller crystallites. The formation of protein crystals and crystal-like ordered aggregates underlies several human pathological conditions. An example is the crystallization of hemoglobin C and the polymerization of hemoglobin S that cause, respectively, the CC and sickle cell diseases [7–10]. The formation of crystals in the eye lens underlies the pathology of several forms of cataract [11, 12]. A unique example of benign protein crystallization in humans and other mammals is the formation of rhombohedral crystals of insulin in the islets of Langerhans in the pancreas [13]. Traditionally, protein crystals have been used for the determination of the atomic structure of protein molecules by X-ray crystallography [14]; this method contributes $\sim$87% of all protein structures solved, with the majority of the other determinations carried out by nuclear magnetic resonance (NMR) spectroscopy [15]. Nanoparticle synthesis can benefit from the advances in these and other research areas.

Below, we first discuss the thermodynamic and kinetic aspects of the classical nucleation theory, which still represents the main framework for the understanding

of nucleation phenomena. Then we consider recent data on the rates of nucleation of protein crystals and show that several of the features of the experimentally determined kinetic dependencies do not comply with the predictions of the classical theory. We then discuss the two-step mechanism of nucleation, according to which the crystalline nuclei appear inside metastable clusters of size several hundred nanometers, which consist of dense liquid and are suspended in the solution. We review recent evidence suggesting that while this mechanism was first proposed for the nucleation of protein crystals, it applies to the nucleation of small-molecule organic and inorganic, as well as colloid and biomineral crystals. We also show that at the high supersaturations employed in many crystallizing systems the nucleation barrier becomes negligible, that is, the generation of the crystals proceeds in the spinodal regime. We discuss the implication of these findings for the nucleation rate, for the nucleation's response to the presence of foreign surfaces, and for the selection of the polymorph form of the crystallizing material.

II. THE CLASSICAL NUCLEATION THEORY

A. The Crystallization Driving Force

In correspondence to the typical physiological, laboratory, and industrial conditions, nucleation is typically considered under constant temperature and pressure. With such constraints, the transfer of solute molecules from solution to the crystal is driven by the change of Gibbs free energy [16]. The change in Gibbs free energy of crystallization, $\Delta G^{\circ}_{\text{cryst}}$, at constant temperature T, is the sum of the contributions of the enthalpy $\Delta H^{\circ}_{\text{cryst}}$ and entropy $\Delta S^{\circ}_{\text{cryst}}$: $\Delta G^{\circ}_{\text{cryst}} = \Delta H^{\circ}_{\text{cryst}} - T\Delta S^{\circ}_{\text{cryst}}$. The associated crystallization equilibrium constant

$$K_{\text{cryst}} \equiv \exp\left(-\Delta G^{\circ}_{\text{cryst}}/RT\right),\ K_{\text{cryst}} = C_{\text{e}}^{-1} \tag{1}$$

where C_{e} is the protein solubility with respect to the studied crystalline form, R is the universal gas constant, and T is the absolute temperature.

Crystal formation occurs in supersaturated solution, in which the concentration C is higher than the solubility C_{e}. Accordingly, the chemical potential of the solute μ in the solution is greater than the one at equilibrium μ_{e}, which in turn is equal to the chemical potential of the crystallizing material in the crystal, $\mu_{\text{e}} = \mu_{\text{crystal}}$. The chemical potential $\mu = \mu_0 + RT \ln \gamma C$ and $\mu_{\text{e}} = \mu_0 + RT \ln \gamma_{\text{e}} C_{\text{e}}$, where γ and γ_{e} are, respectively, the activity coefficients of the solute in the crystallizing solution and in a solution with equilibrium concentration C_{e}, and μ_0 is the chemical potential in a standard solution. Then the nucleation driving force $\Delta\mu = \mu - \mu_{\text{e}} = RT \ln(\gamma C/\gamma_{\text{e}} C_{\text{e}})$. Often, it is assumed that $\gamma = \gamma_{\text{e}}$ so that $\Delta\mu = RT \ln(C/C_{\text{e}})$.

Since γ is a function of concentration, if $C \gg C_{\text{e}}$, the assumption $\gamma = \gamma_{\text{e}}$ is unjustified. In protein and colloid solutions, the activity coefficients are evaluated

from the relation $\ln \gamma \cong 2B_2C$, where B_2 is the second osmotic virial coefficient. B_2 depends in the intermolecular interactions between the solute molecules and can be independently determined from the dependence of the osmotic compressibility $(d\Pi/dC) = RT(KC/R_\theta)$ on the concentration C, where Π is the osmotic pressure, K is an instrument constant, and R_θ is the Raleigh ratio of intensity scattered at angle θ to the incident light intensity; the dependence $KC/R_\theta(C)$ is measured by static light scattering [17, 18].

In solutions of biominerals and other complex salts the supersaturation is defined as $\Delta\mu = RT \ln(a_+a_-/K_{sp})$, where the solubility product $K_{sp} = a_+^e a_-^e$ takes the role of K_{cryst}. The activities of the cations a_+ and anions a_- in the growth solution and at equilibrium, denoted with superscript e, are calculated from the concentrations of the respective species, accounting for the other solution components [19]. For compounds more complex than the binary salts assumed in the above expressions for $\Delta\mu$ and K_{sp}, they become correspondingly more complex.

This definition of $\Delta\mu = \mu - \mu_e = \mu - \mu_{crystal}$, accepted in the fields of phase transformations, nucleation and crystal growth, contradicts the standard definition of the change of a thermodynamic variable in a physical or chemical process. In the standard definition, Δ signifies the difference between the final and initial states, while in the above definition, the crystal is the final state and the solution—the initial. Hence, the two definitions of $\Delta\mu$ differ in sign.

B. The Thermodynamic Theory of J.W. Gibbs

The formation of crystals is a first-order phase transition. Accordingly, it is characterized with nonzero latent heat, the crystallization enthalpy ΔH°_{cryst} discussed above. More significant for the kinetics of nucleation is the second feature of first-order phase transitions: the discontinuity of the concentration at the phase boundary. As a result of this discontinuity, the solution–crystal boundary possesses nonzero surface free energy. If a small piece of a condensed phase forms in a supersaturated solution, the surface free energy of the emerging phase boundary makes this process unfavorable. Thus, a very limited number of embryos of the condensed phase appear as a result of the few fluctuations that overcome the free energy barrier. The first step in the formation of a new phase, in which the kinetics of the phase transformation is determined by this barrier, is called nucleation.

The thermodynamic part of the classical nucleation theory was developed by J.W. Gibbs in two papers [20, 21]. We present it here with a slight modification: we consider the free energy balance of creating a cluster consisting of n molecules of size a, instead of a cluster of radius r, as in the Gibbs's papers. In a supersaturated solution, that is, one in which the solute chemical potential is higher than that of molecules in the crystal so that $\Delta\mu > 0$, the formation of such a cluster leads to a free energy loss of $-n\Delta\mu$. On the other hand, the creation of the phase boundary with area S and surface free energy α between the cluster and the solution leads

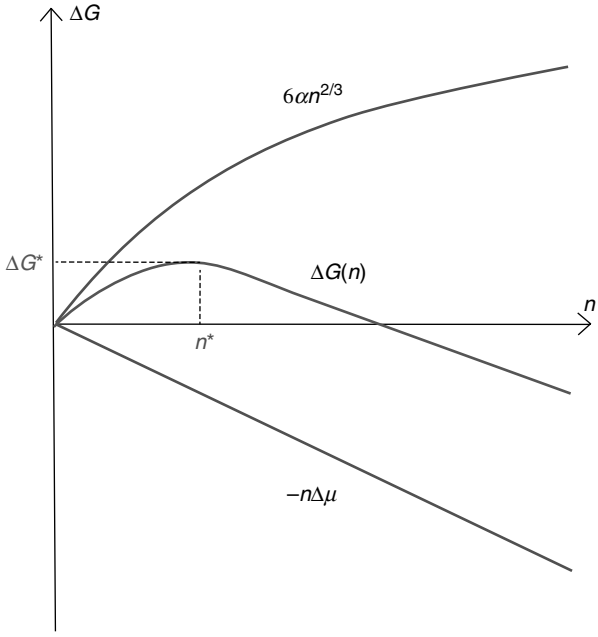


Figure 1. Illustration of the thermodynamic effects of formation of a crystal. n, number of molecules in crystal; $\Delta\mu$, solution supersaturation; α, surface free energy; ΔG, free energy;* denotes critical cluster.

to a free energy gain $S\alpha$. Assuming that the crystal cluster is a cube, $S = 6a^2n^{2/3}$; other shapes will lead to coefficients different than $6a^2$ in this relation, but the 2/3 scaling with n will be preserved for all three-dimensional nuclei. Thus,

$$\Delta G(n) = -n\Delta\mu + 6a^2n^{2/3}\alpha \tag{2}$$

This dependence is plotted in Fig. 1.

Differentiating $\Delta G(n)$, we find the cluster size n^* for which ΔG passes through a maximum ΔG^*

$$n^* = \frac{64\Omega^2\alpha^3}{\Delta\mu^3} \quad \text{and} \quad \Delta G^* = \frac{32\Omega^2\alpha^3}{\Delta\mu^2} = \frac{1}{2}n^*\Delta\mu \tag{3}$$

where $\Omega = a^3$ is the volume occupied by a molecule in the crystal.

The main assumptions behind Eq. (2) are that (i) the nucleus size changes continuously, and that (ii) the surface free energy α does not depend on the nucleus size. Clearly, assumption (i) constrains the applicability of the classical theory to large nuclei, for which describing the addition of or detachment of a new molecule with a continuous variable leads to a relatively small error. For nuclei consisting of less than 100 molecules, and in particular those with fewer than 10 molecules, the application of Eq. (2) and its corollaries may lead to significant deviations

between theoretical predictions and experimental facts [22, 23]. Assumption (ii) is often read to imply that the surface free energy between the nucleus and the solution is the same as the surface free energy of a macroscopic crystal and the same solution. Since determinations of the surface free energy of macroscopic crystals is nearly impossible [24], one cannot claim that the predictive power of Eq. (2) is limited by this version of assumption (ii). The other interpretation of assumption (ii) appears more dangerous: small nuclei of size n would clearly have a different surface structure than small nuclei of size $n+1$. Thus, for small nuclei α would be a function of the nucleus size and this would lead to a different functional form of $\Delta G(n)$ and different location of its maximum in Eq. (3).

There are numerous treatments in which the dependencies $\Delta G(n)$ and $n^*(\Delta\mu)$ have been derived or computed without the help of these two assumptions [23, 25, 26]. Importantly, while they lead to different functional forms of the $\Delta G(n)$ and $n^*(\Delta\mu)$, the main features of Eqs. (2) and (3) are preserved: $\Delta G(n)$ passes through a maximum, from which n^* is defined, and n^* is roughly inversely proportional to $\Delta\mu^3$.

ΔG^* from Eq. (3) is the barrier that must be overcome to form a crystal from solute molecules. The growth of clusters smaller than n^* is associated with an increase of free energy and is unfavorable. Clusters may still grow to such sizes as a result of a fluctuation, but since a driving force exists for the decay of these clusters, such events are rare. On the other hand, if as a result of a fluctuation a cluster reaches a size greater than n^*, its growth is accompanied by a decrease of free energy and occurs spontaneously. A cluster of size n^* has equal probabilities of growth and decay and, hence, such clusters are called critical and they represent the nuclei of the new phase. Note that by this definition all nuclei are critical and the term "critical nuclei" is redundant [25].

C. The Rate of Crystal Nucleation

To model the nucleation rate J, that is, the number of nuclei that appear in a unit solution volume per unit time, M. Volmer postulated—in analogy to the Arrhenius equation—that $J = J_0 \exp(-\Delta G^*/k_B T)$, where k_B is the Boltzmann constant [27]. The external parameters, such as temperature, concentration, and pressure, as well the solution supersaturation, affect the nucleation rate through ΔG^* according to Eq. (3). There are numerous statistical–mechanical derivations of the nucleation rate law within the assumption of the classical nucleation theory, for an example, see Ref. 26. The final expression of these derivations can be represented as [28]

$$J = \nu^* Z n \exp(-\Delta G^*/k_B T) \tag{4}$$

where ν^* is the rate of attachment of monomers to the nucleus, Z is the Zeldovich factor, which accounts for the width of the free energy profile $\Delta G(n)$ in the vicinity of the maximum ΔG^*, see Fig. 1, and n is the number density of molecules in

the solution. Equation (4) assumes that the replacement partition function of the nucleus [26, 28] is equal to 1. This factor accounts for the additional stabilization of the nuclei due to their translational and rotational degrees of freedom [29]. Neglecting it is a reasonable assumption for crystal nuclei suspended in a viscous solution; this would not be the case for nucleation in the gas phase.

A major assumption in the derivation of Eq. (4) is that the solution molecules exchange directly with an ordered cluster. To understand the meaning of this assumption and why it might not apply to nucleation of crystals in solution, we need to step back and consider the distinction between a solution and a crystal.

Let us consider the phase diagram of a solution or any other two-component system in coordinates concentration and temperature at constant pressure. This phase diagram typically contains three phases: a dilute solution, a dense liquid, and crystal; a higher number of phases are possible if more than one crystalline polymorph may form. While with some solutions of small-molecule compounds the dense liquid might not be observable because it would occur at temperatures lower than the solvent freezing point, the dense liquid is readily seen in protein, colloid, and some organic solutions [30–33]. To distinguish between the three phases present in the phase diagram, at least two parameters, called order parameters, are needed. Thus, the dilute solution and the dense liquid differ by the solute concentration, the dense liquid and the crystal differ by structure (there may be a slight difference in concentration), and the dilute solution and the crystal differ by both concentration and structure.

From this point of view, the formation of crystals in solution should be viewed as a transition along two order parameters: concentration and structure [18]. While a fluctuation along the concentration axis is easy to imagine, structure transitions appear less trivial. Pure structure transitions are only possible in melts, whose concentration is similar to that of the emerging crystalline phase. Crystalline nuclei form as a result of a fluctuation along the structure axis. The smallest structure fluctuation can be viewed as a pair of molecules from the melt that has an orientation identical to the orientation of a pair of molecules in the crystal [34, 35]. This crystal-like orientation in the pair is preserved over times significantly longer than the lifetime of a "bond" in the melt. A nucleus arises as a result of accumulation of such ordered pairs into an ordered piece of new phase. In a sense, structure fluctuations can be viewed as fluctuations of the density of ordered pairs.

If a crystal nucleates not from its melt, but from a dilute solution or gas, both a concentration and a structure fluctuation are needed so that a crystalline nucleus may form (Fig. 2a). Thus, the above assumption that an ordered nucleus forms directly in the dilute solution corresponds to the assumption that the solution to crystal transformation occurs as a transition along both order parameters, density and crystallinity, simultaneously; in Fig. 2a this pathway is represented by the arrow along the diagonal of the (concentrations, structure) plane. It could be argued that a more logical pathway for the transition is to proceed along the two order parameters

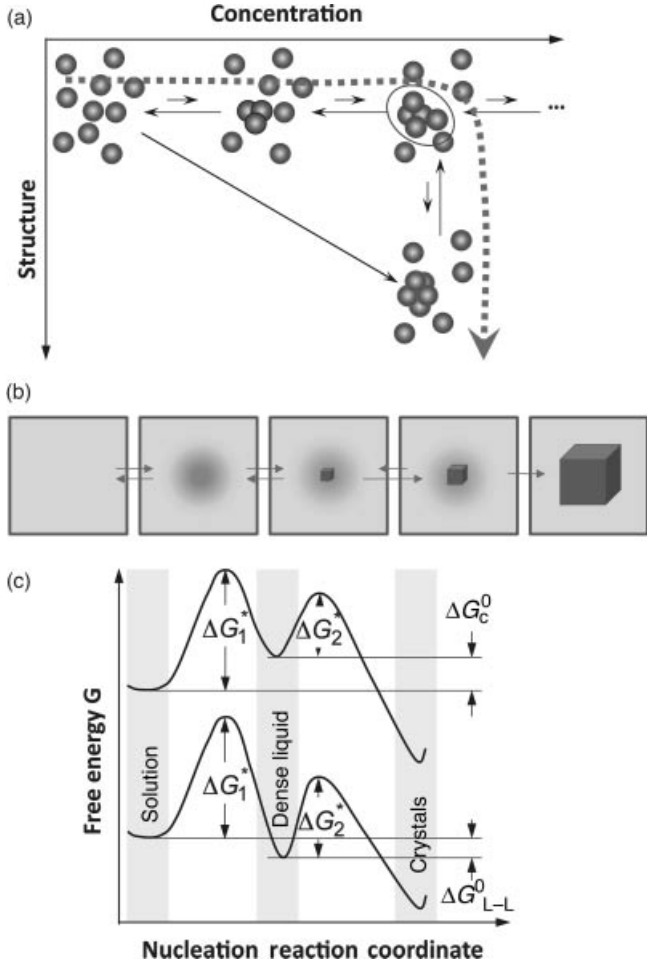


Figure 2. Schematic illustration of the two-step mechanism of nucleation of crystals. A dense liquid cluster forms. A crystal nucleus may form inside the cluster. (**a**) Microscopic viewpoint in the (concentration, structure) plane; (**b**) macroscopic viewpoint of events along dashed line in (a). (**c**) The free energy ΔG along two possible pathways for nucleation of crystals from solution. If dense liquid is unstable and $\Delta G^0_{L-L} > 0$ (ΔG^0_{L-L}—standard free energy of formation of dense liquid phase), dense liquid exists as mesoscopic clusters, ΔG^0_{L-L} transforms to ΔG^0_C, and upper curve applies; if dense liquid is stable, $\Delta G^0_{L-L} < 0$, reflected by lower curve. ΔG^*_1 is the barrier for formation of a cluster of dense liquid, ΔG^*_2—for a structure fluctuation leading to an ordered cluster.

in sequence. Such a sequential pathway would correspond to the formation of droplet of a dense liquid followed by the formation of a crystalline nucleus inside this droplet, as illustrated in Fig. 2b.

This mechanism was first suggested by simulations and analytical theory [36–38]. These theoretical efforts predicted that the density and structure fluctuations are only separated near the critical point for liquid–liquid (L–L) separation occurring in model protein solution systems [36, 39, 40], while for off-critical compositions, the fluctuations of the density and structure order parameters occur synchronously [36], similarly to the classical viewpoint.

The experiments discussed below demonstrate that nucleation of crystals of the protein lysozyme, under a broad range of conditions, proceeds in two steps: the formation of a droplet of a dense liquid, followed by nucleating a periodic crystal within the droplet [41–44], as schematically illustrated in Fig. 2. If the dense liquid is stable with respect to the dilute solution, the nucleation of crystals occurs inside macroscopic droplets of this phase. A far more common case is when the dense liquid is not stable but has a higher free energy than the dilute solution [31, 32]. In these cases, the dense liquid is contained in metastable clusters, intriguing objects in their own right, and crystal nucleation occurs within the clusters.

After and concurrently with the evidence for the operability of the two-step mechanism in the case of lysozyme crystallization, additional experimental results demonstrated that this mechanism applies to many other proteins, to small molecule organic and inorganic compounds, including biominerals, and colloids. Below, we discuss these and other issues related to the two-step nucleation mechanism.

III. THE TWO-STEP MECHANISM AND THE SOLUTION–CRYSTAL SPINODAL

A. Experimental Data on the Rate of Nucleation of Crystals

To understand the mechanism of nucleation of crystals in solution we turn to data on the dependence of the nucleation rate on supersaturation for crystals of the protein lysozyme, a convenient and often used model system. The dependencies of the homogeneous nucleation rate of lysozyme crystals on the thermodynamic supersaturation $\sigma \equiv \Delta\mu/k_{\mathrm{B}}T$ at three different concentrations of the precipitant, NaCl, are presented in Fig. 3. The data in Fig. 3 were obtained using the technique for direct determination of the nucleation rates of proteins discussed in Refs 45, 46, which allows distinction between homogenously and heterogeneously nucleated crystals so that the data points in Fig. 3 are homogenous nucleation rates. In support of the conclusion that the rates plotted in Fig. 3 characterize homogeneous nucleation is the fact that they are lower by several orders of magnitude than those stemming from less careful measurements that may have been contaminated by heterogeneous nucleation events [46–49].

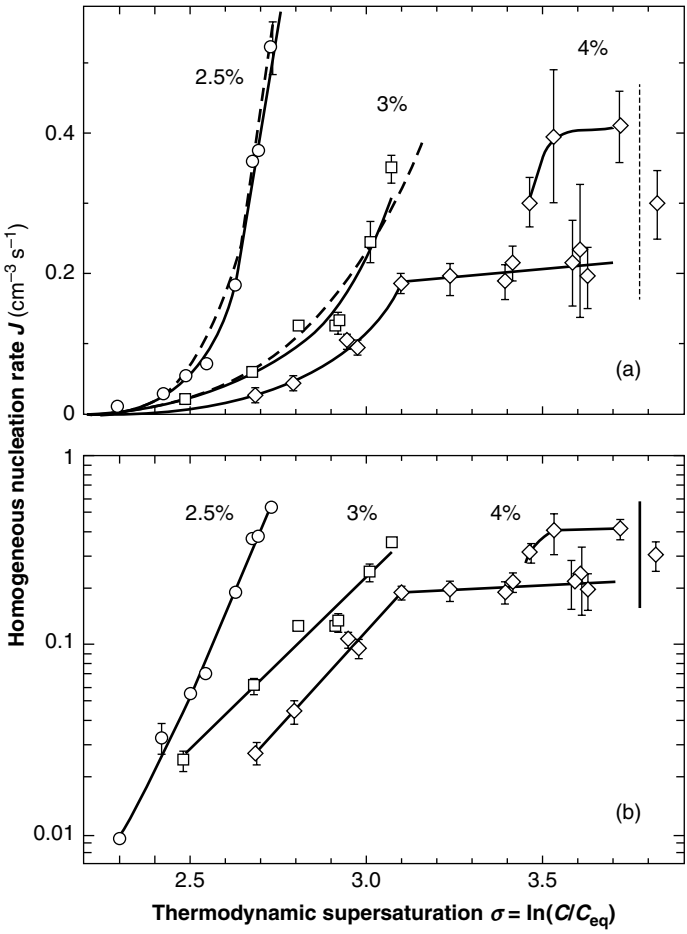


Figure 3. The dependence of the rate of homogeneous nucleation J of lysozyme crystals on supersaturation $\sigma \equiv \Delta\mu/k_BT$ at $T = 12.6^\circ$C and at the three concentrations of the precipitant NaCl indicated on the plots. Solid lines: fits with exponential functions; dashed lines: fits with the classical nucleation theory expression, Eq. (4). Vertical dotted lines at $\sigma = 3.8$ indicate the liquid–liquid coexistence boundary at this T and $C_{NaCl} = 4\%$; this supersaturation corresponds to lysozyme concentration 67 mg mL^{-1}. (**a**) Linear coordinates; (**b**) semilogarithmic coordinates. With permission from Ref. 50.

Each data series in Fig. 3 corresponds to nucleation experiments carried out at a fixed precipitant concentration and at fixed temperature. In agreement with general expectations, the nucleation rate increases exponentially with supersaturation at each of the three precipitant concentrations, and, overall, is higher at higher precipitant concentrations. However, the dependencies contain four peculiarities.

(i) The $J(\sigma)$ dependence at the highest precipitant concentration, $C_{\mathrm{NaCl}} = 4\%$, breaks at $\sigma > 3.1$ and, in dramatic contrast to prediction of Eqs. (3) and (4), the section above this concentration is practically steady as supersaturation increases.

(ii) At $\sigma > 3.45$ in the same $J(\sigma)$ dependence, the data scatter increases and three of the recorded points deviate significantly from the dominant trend.

(iii) The measured nucleation rates are of order 0.1–1 $\mathrm{cm}^{-3}\ \mathrm{s}^{-1}$, which is about 10 orders of magnitude less than the prediction of the classical nucleation theory; the estimate of J stemming from the classical nucleation theory is discussed below.

(iv) The dependence of the nucleation rate on temperature, shown in Fig. 4 presents another puzzling complexity: as supersaturation is increased upon lowering of temperature, the nucleation rate first increases exponentially, as expected from the classical theory, but then passes through as sharp maximum and recedes following a weaker dependence.

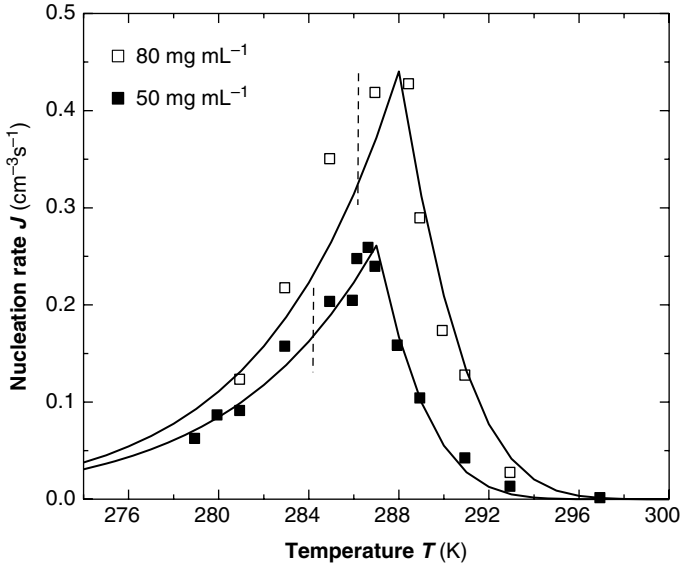


Figure 4. The dependence of the rate of homogeneous nucleation J of lysozyme crystals on temperature T at two fixed lysozyme concentration indicated in the plot. The temperatures of equilibrium between crystals and solution are 315K at $C_{\mathrm{lys}} = 50\,\mathrm{mg\ mL}^{-1}$ and 319K at $C_{\mathrm{lys}} = 80\,\mathrm{mg\ mL}^{-1}$. The temperatures of L–L separation are 285K at $C_{\mathrm{lys}} = 50\,\mathrm{mg\ mL}^{-1}$ and 287K at $C_{\mathrm{lys}} = 80\,\mathrm{mg\ mL}^{-1}$ [32] and are marked with vertical dashed lines. Symbols represent experimental results from [43]. Lines are results of Eqs. (6)–(8). With permission from Ref. 51.

In the following sections, we address these four peculiarities and use them to draw conclusions on features of the nucleation mechanism that go beyond the classical nucleation theory.

B. The Nucleus Size and Solution-to-Crystal Spinodal

To understand the breaking $J(C)$ dependency, feature (i) above, we use the nucleation theorem to determine the size of the critical nucleus for crystallization. According to Eq. (3), the number of molecules in the nucleus n^* largely determines the height of the free energy barrier for nucleation ΔG^*, and hence the nucleation rate J. To determine n^*, we employ the nucleation theorem [52–55], a universal, model-independent nucleation law. Importantly, the nucleation theorem does not rely on the assumptions on the surface free energy of the nucleus, its continuous growth, or the way, in which molecules join the nucleus, discussed above in relation to Eqs. (2)–(4). The nucleation theorem provides an estimate for n^* from the nucleation rate J,

$$n^* - n_0 = k_B T \frac{\partial \ln J}{\partial \Delta \mu} + \alpha_1 \tag{5}$$

where α_1 is a correction that takes values between 0 and 1 [53].

Figure 3b indicates that at $C_{NaCl} = 2.5$ and 3%, n^* does not change throughout the respective supersaturation ranges, while at $C_{NaCl} = 4\%$ the nucleus size changes abruptly at $\sigma = 3.1$, corresponding to $C = 33.5\,\text{mg mL}^{-1}$. The value of the parameter n_0, which roughly corresponds to the number of solution protein molecules displaced by the nucleus, can be roughly estimated as less than 1. Then the nucleus sizes $n^* - n_0$, extracted from the four linear segments in Fig. 3b are 10, 4, 5, and 1 molecules, respectively. From here we see that the breaking in the $J(C)$ dependence at $C_{NaCl} = 4\%$ is due to the transition of the nucleus size from five to one molecules.

Nucleus size $n^* - n_0 = 1$ means that every molecule in the solution can be an embryo of the crystalline phase, and the growth to dimer and larger clusters occurs with a free energy *gain*. Thus, the free energy barrier for the formation of the crystalline phase ΔG^* is below the thermal energy of the molecules. In analogy to the nucleation of a fluid within another fluid, we call *spinodal* the phase line at which the nucleation barrier vanishes and the rate of generation of the new phase is only limited by the kinetics of growth of its clusters. The spinodal is defined as the boundary between metastability and instability of an "old" phase, supersaturated with respect to a "new" phase [20, 21, 56].

The case discussed here, the solution-to-solid phase transition, is one for which a mean-field free energy expression encompassing both phases cannot be formulated because of different standard states. Since the inflection point in the dependence of ΔG on the order parameter along which the phase transition occurs is typically

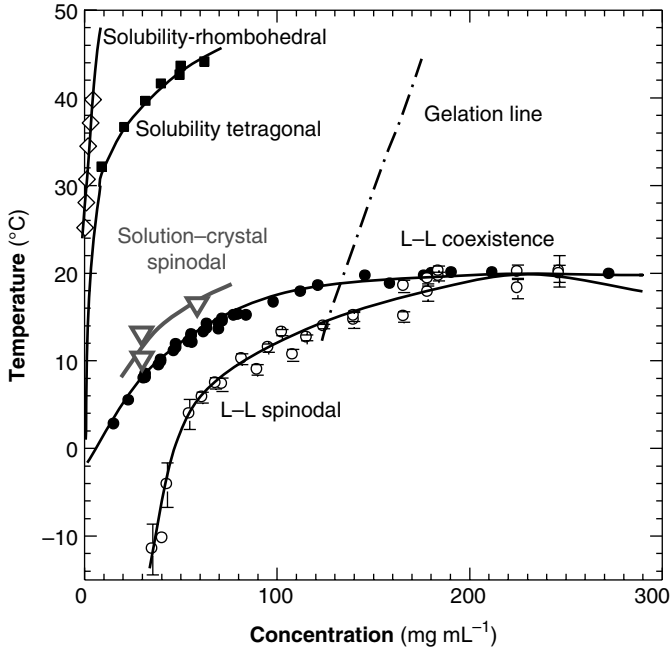


Figure 5. The phase diagram of a lysozyme solution determined experimentally in 0.05 M Na acetate buffer at pH = 4.5 and 4.0% NaCl. Liquidus, or solubility lines [61, 62], liquid–liquid (L–L) coexistence and respective spinodal [32], gelation line [31, 32]. Solution-to-crystal spinodal is highlighted in grey and is from Ref. 63.

used to define the spinodal [57–59], a thermodynamic definition of the solution-to-crystal spinodal is impossible [57]. The definition proposed here is a kinetic one, based on the transition to nucleus size of *one* molecule, that is, to where no thermodynamic barriers for the formation of the crystalline phase exist.

In Fig. 5, we have depicted the solution-to-crystal spinodal line in the (C,T) plane, determined as the concentration C at the transition to $n^* - n_0 = 1$ from Ref. 60. Since at concentrations and temperature below this spinodal line $\Delta G^* \approx 0$, the nucleation rate J does not increase as supersaturation is increased by increasing C or lowering T. This explains puzzle (i) above. The existence of a solution-to-crystal spinodal also helps to explain the maxima in the dependencies of the nucleation rate J on temperature in Fig. 4, puzzle (iv) above; for further details and a theoretical model of these factors, see below.

The transition to a spinodal regime of crystal formation also explains the increased data scatter of $J(\sigma)$ at $\sigma > 3.45$, puzzle (ii) above. As shown in Refs 43, 51, at the point of transition from nucleation to spinodal decomposition the nucleation

rate undergoes a sharp maximum: on the one side is an ascending branch due to the decrease of the size of the nucleus, and on the other side is a descending branch due to the temperature decrease and associated kinetic factors. Near this maximum, the nucleation rate is very sensitive to variations of the experimental conditions: temperature, protein and precipitant concentrations, and others. Hence, minor inconsistencies of these parameters may lead to significant variations in J [42].

C. The Classical Theory Overestimates the Crystal Nucleation Rate by 10 Orders of Magnitude

To understand puzzle (iii) above, we use Eq. (4) for an estimate of the crystal nucleation rate based on the classical nucleation theory. The rate ν^* can be evaluated from the rate of attachment of molecules to lysozyme crystals at similar protein concentrations. As discussed in Ref. 64, the surfaces of crystal growing in solution are smooth and molecules only attach to growth steps which occupy about 10^{-3}–10^{-2} of the crystal surface. Hence, the rate of attachment to crystals should be estimated from the velocity of step propagation rather than from the rate of growth of the crystal faces.

There are numerous determinations of the step velocities of lysozyme crystals [65–67]. At temperatures and concentrations similar to those during the determination of the nucleation rate in Fig. 3, the step velocities are $\sim 1\ \mu m\ s^{-1}$. This yields, with molecular size of lysozyme of 3.5 nm, attachment rate to the steps $\sim 300\ s^{-1}$. In contrast to that of large crystals, the nucleus surface is likely rough (because of the small size of the nucleus) and molecules can attach anywhere. Hence, we assume that $\nu^* \approx 300\ s^{-1}$.

The Zeldovich factor Z accounts for the width of the free energy profile along the nucleation reaction coordinate around the location of the maximum [25, 26, 34, 35]. It is expected to be of order 0.1–0.01 for nucleation of any protein condensed phase [25, 35, 50]. The protein number density in a solution of concentration ~ 50 mg mL^{-1} as the one used for the experiments in Fig. 3 [68] is $n = 2 \times 10^{18}\ cm^{-3}$. With these values for ν^*, Z, and n, the pre-exponential factor in Eq. (4) is of order 10^{19}–$10^{20}\ cm^{-3}\ s^{-1}$.

The nucleation barrier ΔG^*, determined from the slope of the dependencies in Fig. 3b $\Delta G^* \approx 10^{-19}$ J. We can use Eq. (3) to evaluate the surface free energy α of the interface between the dense liquid and the solution from the value of ΔG^*. From the crystal structure, $\Omega \cong 3 \times 10^{-20}\ cm^3$ [69]. We get $\alpha = 0.5$–0.6 erg cm^{-2} [50], which is close to determinations for number of other protein crystals [70, 71] and this correspondence supports the estimate of ΔG^* from the data in Fig. 3.

Combining the estimate for the pre-exponential factor with this estimate for ΔG^* from Eq. (4) we get a prediction for $J \approx 10^8$–$10^9\ cm^{-3}\ s^{-1}$. This value is

about 10 orders of magnitude higher than the one in Fig. 3. It is important to note that since we estimate ΔG^* from experimental data, the difference between the experimentally determined J and the prediction of the classical nucleation theory is due to a lower pre-exponential factor.

D. The Two-Step Mechanism of Nucleation of Crystal in Solution

To understand puzzles (iii) and (iv) above, that the nucleation rate is lower by many orders of magnitude than the prediction of the classical theory and the non-monotonic dependence of the nucleation rate on temperature, we show below that the nucleation of crystals occurs inside metastable mesoscopic clusters of dense protein liquid, as illustrated in Fig. 2.

Direct observations of ordered nuclei forming within the dense liquid exist, but only for the case of stable dense protein liquid, Figs. 6 and 7 [72, 73]. Such direct imaging would be difficult or impossible for the more common case in which the dense liquid is unstable. The action of the two-step mechanism in this case is inferred from two pieces of evidence: First, we demonstrate the existence metastable mesoscopic dense liquid clusters in solutions. Then, we analyze the complex kinetic curves in Figs. 3 and 4, propose a kinetic law for the two-step mechanism and show that its predictions qualitatively and quantitatively agree with the experimental data.

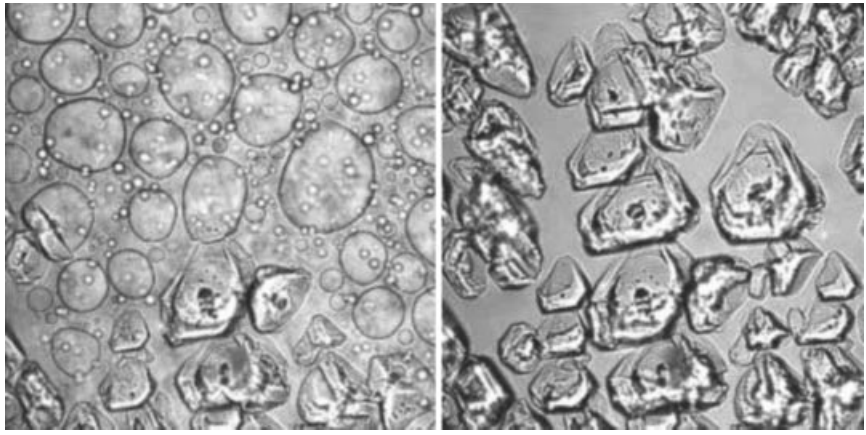


Figure 6. Confocal scanning laser fluorescence microscopy imaging of nucleation of crystals of glucose isomerase within dense liquid droplets. Bright field imaging; polyethylene glycol with molecules mass 10,000 g mol^{-1} (PEG 10000) used to induce crystallization. The time interval between the left and right images is 380 s. $C_{\text{protein}} = 55$ mg mL^{-1}, $C_{\text{PEG}} = 9.5\%$, 0.5 M NaCl, 10 mM Tris pH 7; dimensions of each image: 326 μm × 326 μm. With permission from Ref. 72.

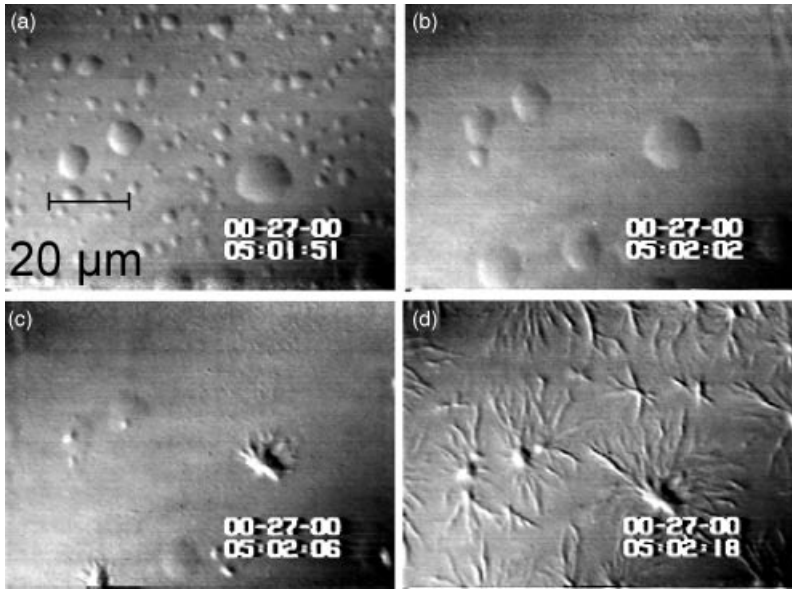


Figure 7. The nucleation of polymers of deoxy sickle cell hemoglobin (HbS) in dense liquid droplets existing in solution of this protein. Concentration of HbS is 220 mg mL^{-1}. (**a**)–(**d**) Temperature is lowered from 42 to 35°C, the smaller of the dense liquid droplets disappear, while the larger ones serve as nucleation centers for HbS spherulites. Spherulites also appear at the locations where smaller droplets have been, apparently because of the undissipated locally higher concentration. Time elapsed between (a) and (d) is 27 s, as indicated in the panels. With permission from Ref. 73.

E. Dense Liquid Clusters

If crystallization is carried out at a point in the phase diagram where the dense liquid is unstable, all density fluctuations are expected to decay with a characteristic time of order of the diffusion time of the protein molecules, 10 μs, see below [74–76]. Since the molecules in the region of high concentration within the fluctuation move with the same characteristic time, it would be impossible for them to probe various structures and find the right one for the crystalline nucleus. Thus, the crucial question for the understating of nucleation from dilute media is: How does the transition along the order parameter concentration occur? The answer lies in the recently discovered metastable mesoscopic clusters of dense liquid.

The evidence for metastable dense liquid clusters comes from monitoring solutions of three hemoglobin variants, oxy-HbA, oxy-HbS, and deoxy-HbS, and the proteins lumazine synthase and lysozyme, by dynamic light scattering (DLS). Figure 8a shows a typical intensity correlation function of a lysozyme solution in the homogeneous regions of the phase diagram. The correlation function reveals

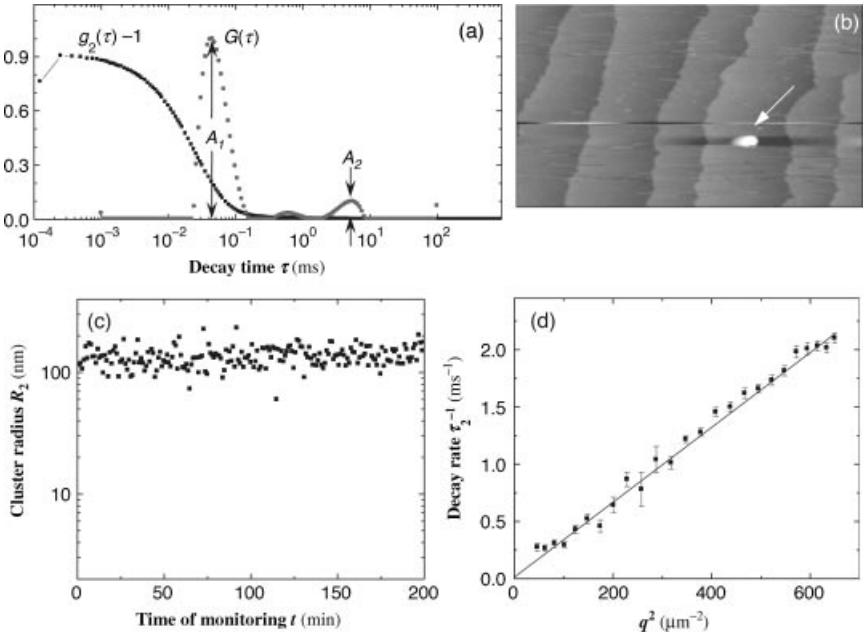


Figure 8. Characterization of dense liquid clusters. (**a**) Examples of correlation function of the scattered intensity $g_2(\tau)$ and the respective intensity distribution function $G(\tau)$ of a lysozyme solution with $C = 148$ mg mL^{-1} in 20 mM HEPES buffer; data collected at angle 145°. (**b**) Atomic force microscopy imaging of liquid cluster landing on the surface of a crystal in a lumazine synthase solution. Tapping mode AFM imaging, scan width 20 μm. Apparent lateral cluster dimensions are misleading, cluster height is 120 nm. (**c**) Time dependence of the radius of dense liquid clusters in the same lysozyme solution as in (a). (**d**) The dependence of the decay rate τ_2^{-1} of the cluster peak in the correlation function on the squared wavevector q^2 for a lysozyme solution as in (a). With permission from Ref. 78.

two processes: the faster process, with characteristic time of ~10–100 μs, is the Brownian motion of single lysozyme molecules; it is present at all solution concentrations. The corresponding hydrodynamic radius, determined via the Stokes–Einstein equation, is about 1.5 nm and matches well the diameter of a lysozyme molecule of 3.2 nm. The slower process takes milliseconds; its amplitude increases with higher lysozyme concentrations. This longer time could come from either compact lysozyme clusters suspended in the lysozyme solution, or from single lysozyme molecules embedded in a loose network structure constraining their free diffusion. Since the measured low shear viscosity of lysozyme solutions is equal to those determined using high shear rates [77], no loose networks in lysozyme molecules exist in these solutions and we conclude that long times in Fig. 8a indeed correspond to lysozyme clusters [74]. The time-dependence of their radius is shown

in Fig. 8c. The clusters appear immediately after solution preparation; their radius is relatively steady (Fig. 8). We therefore conclude that these are clusters of dense liquid.

The number density n_2 of the dense liquid clusters and the fraction of the total solution volume φ_2 they occupy are evaluated from the amplitudes A_1 and A_2 of the respective peaks in the distribution function [74]. Further results on the behavior of clusters of dense liquid in solutions of hemoglobin and lumazine synthase are presented in Refs 74, 76, 78. It was found that with all studied proteins, the clusters exist in broad temperature and protein concentration ranges. The clusters occupy $\varphi = 10^{-6} - 10^{-3}$ of the solution volume [74].

To evaluate the lifetime of the lysozyme clusters, we note that cluster decay processes contribute a q-independent component to the overall rate sensed by DLS Ref. 79, $\Gamma_2 = \Gamma_0 + D_2 q^2$, and can be distinguished from cluster diffusion. (Γ_0 is the rate of cluster decay, D_2 is the cluster diffusion coefficient, and q is the wavevector.) The q-dependent, diffusion component indeed dominates the DLS signal (Fig. 8d). Using $\Gamma_0 \ll D_2 q^2$ with $q^2 = 3.5 \times 10^{10}\,\mathrm{cm}^{-2}$ and $D_2 = 2 \times 10^{-9}\,\mathrm{cm}^2\ \mathrm{s}^{-1}$, $\Gamma_0 \ll 70\,\mathrm{s}^{-1}$, we obtain a lower bound $1/\Gamma_0 \approx 15$ ms for cluster lifetimes.

The determination of the lifetime of the clusters of lumazine synthase was more straightforward and yielded an estimated of $\sim$10 s [76, 78]. In addition to detection by dynamic light scattering, clusters of lumazine synthase were directly imaged by atomic force microscopy (Fig. 8b) [76, 78], which confirmed their macroscopic lifetimes.

The lifetimes of the clusters ($>$15 ms for Hb and lysozyme and $\sim$10 s for lumazine synthase) significantly exceed the equilibration times of the protein concentration at submicrometer length scales, that is, $\sim 10^{-5}$ s. Thus, the compact clusters represent a *metastable* phase separated from the bulk, dilute solution by a free energy barrier.

Attempts to rationalize the finite size of clusters have focused on a balance of short-range attraction, due to van der Waals, hydrophobic or other forces, and screened Coulombic repulsion between like-charged species [80, 81]. While small clusters that contain about 10 particles, naturally appear in such approaches, large clusters are expected only if the constituent particles are highly charged, with about hundreds elementary charges. Such high charges are feasible for micron-size colloidal particles; however, proteins in solution are known to carry less than 10 elementary changes per molecule. Hence, while for colloidal suspensions these theories successfully predict aggregation [82–84], or even the existence of metastable clusters [85], we conclude that a distinct mechanism is at work in protein systems, where clusters contain as many as 10^6 molecules [75]. A recent study concluded that the clusters consist of a nonequilibrium mixture of single protein molecules and long lived but ultimately unstable complexes of proteins [75]. The puzzling mesoscopic size of the clusters is determined by the lifetime and diffusivity of these complexes. Several possible mechanisms of complex formation: domain

swapping, hydration forces, dispersive interactions, and other, system-specific interactions were highlighted.

F. The Rate Law for the Two-Step Mechanism of Crystal Nucleation

A phenomenological theory was developed that takes into account intermediate high-density metastable states in the nucleation process [51]. The rate law for the dependence of the nucleation rate on protein concentration and temperature emerging from this theory is

$$J = \frac{k_2 C_1 T \exp(-(\Delta G_2^*/k_B T))}{\eta(C_1, T)[1 + (U_1/U_0)\exp(\Delta G_C^0/k_B T)]} \tag{6}$$

where the constant k_2 scales the nucleation rate of crystal inside the clusters, C_1 is the protein concentration inside the clusters, that is, $\sim$300 mg mL^{-1}, ΔG_2^* is the barrier for nucleation of crystals inside the clusters, η is the viscosity inside the clusters, U_1 and U_0 are the effective rates of, respectively, decay and formation of clusters at temperature T, and ΔG_C^0 is the standard free energy of a protein molecule inside the clusters in excess to that in the solution, depicted schematically in Fig. 2c [51]. Recent experimental determinations indicate that ΔG_C^0 is of order 10 $k_B T$ [75].

Following Ref. 86, the nucleation barrier ΔG_2^* in the vicinity of the solution-to-crystal spinodal was modeled as

$$\Delta G_2^*(T) = \frac{E^*}{(T_e - T)}\left[1 - \frac{(T_e - T)^2}{\left(T_e - T_{sp}\right)^2}\right] \tag{7}$$

where E^* is a parameter, T_e is the temperature, at which a solution of the studied concentration is in equilibrium with a crystal, and T_{sp} is the spinodal temperature. T_e and T_{sp} are determined from the phase diagram in Fig. 5, and E^* is determined by fitting Eq. (7) to the slope of the $J(C)$ dependencies in Fig. 3b.

The viscosity inside the dense liquid clusters was modeled as

$$\eta = \eta_0 \left\{1 + [\eta] C_1 \exp\left(k_\eta [\eta] C_1\right)\right\} \exp\left(-E_\eta/k_B T\right) \tag{8}$$

where $[\eta]$ is the viscosity increment, and k_η and E_η are constants; all three viscosity parameters are determined from the known dependencies of viscosity in the studied solution on temperature and concentration.

A crucial assumption in Eq. (8) is that the concentration inside the dense liquid clusters C_1 increases as temperature is lowered, in agreement with the phase diagram in Fig. 5 and the likely similarity between the dense liquid in the clusters and the stable dense liquid depicted in the phase diagram [51]. As a result of this $C_1(T)$ dependence, the viscosity η increases much more strongly in response

to decreasing temperature T then suggested by the quasi-Arrhenius member of Eq. (8) with E_η about 10–20 kJ mol^{-1} [87].

The denominator of Eq. (6) offers another pathway by which decreasing temperature affects the nucleation rate J, besides the temperature dependence of the viscosity. Since $(U_0/U_1)\exp(-\Delta G_C^0/k_B T)$ is the nonequilibrium volume fraction occupied by the clusters, the term in the square brackets in the denominator of Eq. (6) is approximately the reciprocal nonequilibrium volume fraction occupied by the clusters, ϕ_2^{-1}. Since $\Delta G_C^0 > 0$, see above, lower T leads to a greater value of the denominator, which corresponds to a lower volume of the dense liquid clusters and accordingly to lower J. This contributes about factor of five in the decrease in J as temperature is lowered from T_{sp} to the lowest values probed in Fig. 4.

Using Eqs. (6)–(8) nucleation rate data at varying temperature and protein concentrations in Fig. 3 and Ref. 63, as well as nonmonotonic dependencies of the nucleation rate on temperature in Fig. 4 were reproduced with high fidelity using literature values or independently determined parameters of the thermodynamic and kinetic parameters of the system [51]. The good correspondence between the model results and the experimental data supports the validity of the two-step nucleation mechanism. According to Eq. (6), the increasing part of the $J(T)$ as temperature is lowered below T_e is due to the increase of the supersaturation $\Delta\mu$ that shrinks ΔG_2^* according to Eq. (3); this leads to exponential increase in the nucleation rate J. The maximum in $J(T)$ is reached exactly at $T = T_{sp}$, where ΔG_2^* vanishes; note that T_{sp} is independently determined from plots similar to the one at 4% in Fig. 3c [63]. The steep decrease in the nucleation rate as T is lowered beyond the maximum at T_{sp} is a crucial part in the proof of the validity of the two-step mechanism: within the two-step mechanism this steep decrease is explained by the smaller volume of the dense liquid clusters at lower temperature, and by the higher concentration inside them, leading to higher viscosity. Both the lower volume of the clusters and the higher viscosity lead to lower nucleation rate.

No pathway of steep decrease of nucleation rate beyond the spinodal temperature exists if one assumes one-step nucleation: nuclei forming within the dilute solution would be exposed to its viscosity, which is a weak function of temperature. Thus, the nucleation rate would decrease almost imperceptibly, by $\sim$16%, assuming $E_\eta = 20$ kJ mol^{-1}, within the 5–6K range probed. Note that the decrease in nucleation rate in glass forming melts in response to temperature decrease, interpreted as a result of viscosity increase in the melt, occurs over 40–50K [88]; furthermore, this response is significantly enhanced by the stronger temperature dependence of viscosity of melts as compared to the viscosity of solutions.

To understand puzzle (iii) above, that the nucleation rate is lower by 10 orders of magnitude than the prediction of the classical theory, we compare the nucleation kinetic law in Eq. (6) to that in Eq. (4). We see that $\phi_2 k_2 C_1 T/\eta$ takes the place of the product νZn. In solutions of concentration C in the range 20–60 mg mL^{-1} as the ones in which the nucleation rates in Fig. 3 were measured, the cluster volume

fraction ϕ_2, represented by the denominator in Eq. (6), is of order 10^{-7}–10^{-6}. With the concentration C_1 in the clusters around 300 mg mL^{-1}, Eq. (8) shows that the viscosity η of the dense liquid in the clusters is around 100 cP, or $\sim 100 \times$ higher than in the normal solution. We get that the nucleation rate should be $\sim 10^9 \times$ lower than the prediction of the classical theory, which assumes nucleation in the solution bulk. Thus, the two-step mechanism explains the third peculiarity of the nucleation rate data in Fig. 3: the significantly lower pre-exponential factor in the nucleation rate law.

G. The Rate-determining Step in the Two-Step Nucleation Mechanism

The derivation of Eq. (6) is based on the assumption that the first step in the two-step mechanism, the formation of the dense liquid clusters is fast and that the second step, the formation of the crystal nuclei within the dense liquid clusters, is rate determining. While the excellent agreement between the experimental data and the prediction of Eq. (6) in Fig. 4 can be viewed as a support of this assumption, it should and can be tested independently.

As first evidence in favor of the fast rate of generation of the dense liquid clusters, we view data on the time dependence of three characteristics of the cluster population: average radius, number density, and volume fraction, illustrated for the case of average cluster radius in Fig. 8c. All of these dependencies, monitored for the proteins lumazine synthase [76, 78], lysozyme [75], and three hemoglobin variants [74] reveal that the clusters appear within several seconds of solution preparation. After that, the cluster populations are stable for several hours.

For an additional test, we use the similarity between the clusters and stable droplets of dense liquid that exist below the liquid–liquid coexistence line in the phase diagram in Fig. 5. The rate of nucleation of the dense droplets was determined by monitoring the increase in time of the number of droplets appearing in an isothermal solution supersaturated with respect to the formation of dense liquid [68]. These data yield droplet nucleation rates, which are of order 10^8 cm^{-3} s^{-1}. These rates are about 10 orders of magnitude faster than the rates of crystal nucleation and support the conclusion that the nucleation of the dense liquid precursors, stable or unstable, is much faster than the rate of crystal nucleation within these precursors.

The conclusion that the rate of nucleation of crystals within the dense liquid clusters is the rate-determining step in the two-step nucleation mechanism supports the applicability of Eq. (6) as the rate law for this process. Another important consequence of this conclusion is related to the applicability of the nucleation theorem to the two-step nucleation mechanism. Since cluster formation is fast, the clusters can be considered in equilibrium with the solution. Then the chemical potential of the protein in the clusters is equal to the chemical potential of the protein in the solution, and $\Delta\mu = \mu_{\text{solute}} - \mu_{\text{crystal}}$ is the supersaturation to which

the crystal nuclei are exposed within the clusters. Since the cluster number is steady, J is the rate of nucleation of crystals inside the clusters. From the latter two conclusions, it follows that applying the nucleation theorem, Eq. (5) with the macroscopically observed nucleation rate and the external supersaturation, is equivalent to applying the nucleation theorem to the nucleation of crystalline the dense liquid. Hence, the size of the nuclei determined using the nucleation theorem refers to the crystalline nuclei within the clusters. Furthermore, the transition to spinodal regime occurs when the crystalline nuclei reach size one molecule and this transition corresponds to $\Delta G_2^* = 0$.

Finally, we can resolve an apparent controversy. From the above estimate of the lowering of the nucleation rate due to the low volume fraction and the high viscosity of the dense liquid, it may appear that the selection of the two-step mechanism violates the principle of fastest increase of entropy [89, 90]. This principle governs the selection of kinetic pathways toward, in most cases, the mechanism leading to the fastest rate: faster consumption of supersaturation corresponds to faster increase of the total entropy of the universe. This is an incorrectly posed problem: the estimate of the nucleation rate above used the value of the nucleation barrier ΔG^* extracted from the experimental data. As just demonstrated, this barrier is in fact ΔG_2^* from Fig. 2c and Eq. (6), that is, the barrier for nucleation of crystals inside the clusters. Since the surface free energy at the interface between the crystal and the solution is likely significantly higher than at the interface between the crystal and the dense liquid, the barrier for nucleation of crystals from the solution would be much higher. This would lead to much slower nucleation of crystals directly from the solution than inside the clusters. Thus, the protein crystal nucleation follows the two-step nucleation mechanism because it provides for faster rate of the solution to crystal phase transition and in this way for faster decrease of the free energy of the system, which corresponds to faster increase of the entropy of the universe.

H. The Role of Heterogeneous Nucleation Substrates

Knowing that the nucleation of crystal within the dense liquid clusters is the rate-limiting step in the two-step mechanism, we can address a broader related question: Since from a general point of view, the rate of nucleation via the two-step mechanism depends on two pre-exponential factors, J_{01} and J_{02}, and two barriers, ΔG_1^* and ΔG_2^*, which of these four parameters is the most significant. Clearly, the answer should be sought between J_{02} and ΔG_2^*. Since nucleation occurs in the vicinity of the solution-to-crystal spinodal, ΔG_2^* is very small, and hence, the most important parameter is J_{02}. This is a surprising conclusion, and it sheds light on the role of heterogeneous substrates in nucleation.

Nucleation is often facilitated by heterogeneous centers [57, 91]. The generally accepted mechanism of heterogeneous nucleation is that it follows the kinetic law

for homogeneous nucleation but is faster due to lowering of the nucleation free energy barrier [57]. Since we now know that ΔG_2^* is insignificant, we conclude that in contrast to the generally accepted viewpoint heterogeneous nucleation centers assist nucleation not by lowering ΔG_2^*, but by assisting the growth of the ordered clusters through the factor accounted for in the pre-exponential factor J_{02}.

There may be many mechanisms by which a surface may facilitate the growth of the ordered clusters. The most obvious one is that the "right" crystal structure, that is, the one that minimizes the free energy of the system, is similar to the structure of the surface. Alternatively, the surface structure may stabilize a necessary intermediary *en route* to the "right" crystal structure, similar to the way enzymes stabilize the transition state, and not the final product of the catalyzed reaction [92]. Another possibility is that the surface may catalyze the formation of the intermolecular bonds in the crystal. If the structure of a substrate is similar to the structure of the growing crystal, this is referred to as templating [93, 94]. Examples were found for crystallization of proteins on mineral substrates and on ordered lipid layers [95, 96]. One may view the acceleration of nucleation of γ-glycine crystals in the bulk of a supersaturated solution by elliptically polarized light, and α-glycine crystals by linearly polarized light as examples of assisted structuring of the dense liquid by appropriately structured electric field [97].

I. The Broad Applicability of the Two-Step Nucleation Mechanism

Above, we analyzed in detail data on the kinetics of nucleation of crystals of the protein lysozyme, which allow a rather confident conclusion about the applicability of the two-step mechanism. The evidence for the applicability of this mechanism to the nucleation of crystals of other proteins is less direct. In Ref. 98, crystals of several intact immunoglobins were found to coexist for extended lengths of time with dense liquid droplets without the droplets generating additional crystal nuclei. The crystals that were nucleated on the droplet boundaries grew into the dilute solution, rather than into the dense liquid. This was interpreted in favor of nucleation of the crystals within dense liquid clusters suspended in the solution.

Besides the nucleation of protein crystals, the action of the two-step mechanism has recently been demonstrated for the homogeneous nucleation of HbS polymers, with metastable dense liquid clusters serving as precursor to ordered nuclei of the HbS polymer [74, 99, 100]. Other studies have shown that the nucleation of amyloid fibrils of several proteins and peptide fragments, such as Alzheimer-causing A-β-peptide or the yeast prion protein follows a variant of the two-step mechanism in which the role of the intermediate liquid state is played by a molten globule of consisting of unfolded protein chains [101, 102].

The applicability of the two-step mechanism to the nucleation of crystals of urea and glycine was deduced in a series of experiments, in which high power laser pulses were shone on supersaturated solutions [97, 103]. It was fond that the

nucleation rate increases as a result of the illumination by 8–9 orders of magnitude and that by using elliptically or linearly polarized light, α- or γ-glycine crystals could be preferentially nucleated. Since glycine does not absorb the illumination wavelength, and the electric field intensity was insufficient to orient single glycine molecules, it was concluded that the elliptically or linearly polarized pulses stabilize the structure fluctuations within the dense liquid, which lead to the respective solid phases [41, 103].

Colloid systems are the ones for which the evidence in favor of the applicability of the two-step mechanism is the strongest. By tracking the motions of individual particles of size a few microns by scanning confocal microscopy, the nucleation of crystals in colloidal solutions was directly observed [104–106]. These experiments revealed that the formation of crystalline nuclei occurs within dense disordered and fluid regions of the solution.

The role of an amorphous precursor in the nucleation of crystal of biominerals has been speculated for a long time, for a historic overview, see [107]. However, it was envisioned that the precursor does not facilitate that formation of the crystalline nuclei, but only serves as a source of material for reprecipitation into a crystalline phase. Only recently it was shown that amorphous or liquid clusters of calcium and carbonate ions are present in calcium carbonate solutions and facilitate the nucleation of calcite crystals, in a manner similar to the role of the mesoscopic clusters in lysozyme crystallization discussed above [108, 109]. The free energy landscape along the nucleation reaction pathway in Fig. 2c was used to characterize kinetics of the process of calcite crystallization [109].

A two-step nucleation mechanism going through metastable clusters (in this case, swollen micelles) has also been theoretically predicted for a ternary system of two homopolymers and their block copolymer [110].

Stable dense liquid was found to exist in solutions of organic materials and serve as location where crystals nucleate and grow [33]. The existence of the dense liquid in these solutions has been attributed to the same fundamental physical mechanism as the one acting in protein solutions: the size of the solute molecules is larger than the characteristic lengthscale of the intermolecular interactions in the solution [40]. On the other hand, unpublished evidence from the pharmaceutical industry suggests that in many other cases the stable dense liquid, referred to as “oil” by the practitioners in the field, is so viscous that no crystals can form in it. This is in contrast to the observations in Figs. 6 and 7, in which crystals and sickle cell hemoglobin polymers form in the relatively nonviscous dense protein liquid. While this has not been tested, it is possible that the two-step mechanism operates in these organic systems by utilizing dense liquid clusters, similar to those seen in protein, colloid, and calcium carbonate solutions.

In general, the two possible intermediate states for the two-step mechanism, the stable dense liquid and the metastable clusters, have distinct mechanisms: the discrepancy of the lengthscale of the intermolecular interactions in the solution

and the size of the crystallizing molecules for the stable dense liquid, and the existence of limited lifetime complexes for the clusters. Thus, for a given system the availability of any of these two intermediate states is independent of the other; both of them depend on the exact physicochemical characteristics of the system.

The broad variety of systems in which the two-step mechanism operates suggests that its selection by the crystallizing systems in preference to the nucleation of ordered phases directly from the low-concentration solution may be based on general physical principles. This idea is supported by two examples of general physical theory: by Sear [111] and by Lutsko and Nicolis [112]. Of particular interest is the latter work. It treated a range of points in the phase diagram of two different model systems that likely encompass a broad variety of real solutions and demonstrated that the two-step formation of crystalline nuclei, via a dense liquid intermediate, encounters a significantly lower barrier than the direct formation of an ordered nucleus and should be faster. Interestingly, the intermediate state resulting from the theory was not stabilized and represents just a well-developed density fluctuation.

IV. SUMMARY AND CONCLUSIONS

We show that the classical nucleation theory fails to provide understanding of several features of measured kinetic curves: nucleation rates, which are orders of magnitude lower than the classical prediction; nucleation kinetics curves that exhibit saturation, or, even more puzzling, maxima and decreasing branches, with increasing supersaturation, as well as the role of the other, stable and unstable, phases possible in solution.

We show that these features of the nucleation kinetics reflect the action of two factors, which are unaccounted by the classical nucleation theory: the existence of a spinodal for the solution to crystal phase transition, and the action of a two-step nucleation mechanism. As the spinodal is reached upon supersaturation increase, the barrier for nucleation of crystals vanishes and further increases in supersaturation do not yield faster nucleation rate. According to the two-step mechanism, the nucleation of crystal, step two, occurs within mesoscopic clusters of dense liquid, step one. While the initial thought provoking results on the nucleation kinetics were obtained for the nucleation of protein crystals, and, correspondingly, the two-step mechanism was first proposed for these types of crystals only, further investigations have shown the validity of this mechanism to organic, inorganic, and colloid materials, including the important class of biominerals.

Since the main body of experimental data supporting the concepts of the solution to crystal spinodal and the two-step mechanism were obtained with protein solutions, a crucial question is the general applicability of these concepts to other crystallizing systems.

The issue of the spinodal appears more straightforward: the nucleation of numerous crystals in industrial and laboratory practice is carried out at such high supersaturations that the nucleation occurs either in the spinodal regime or in the immediate vicinity of this regime, where the nucleus consist of just a few molecules.

The action of the two-step mechanism relies on the availability of disordered liquid or amorphous metastable clusters in the homogeneous solutions prior to nucleation. While such clusters have been demonstrated for several protein systems and for calcium carbonate solutions it is likely that not all solutions would support the existence of such clusters with properties allowing the nucleation of crystals in them. In such systems the action of the direct nucleation mechanism might be the only option. On the other hand, an intriguing hypothesis is presented by one of the theories discussed above: that a stabilized intermediate state, as a stable dense liquid, as seen in Figs. 6 and 7, or as a metastable mesoscopic cluster, as in Fig. 8, is not needed and the two-step mechanism will act even if the intermediate step is just a density fluctuation. Thus, the two-step mechanism may in fact operate in systems where no intermediate is independently found.

Acknowledgments

The ideas discussed in this perspective paper have crystallized over the years in engaging discussions with many colleagues. An incomplete list includes A. Kolomeisky, V Lubchenko, A. Myerson, B. Garetz, D. Oxtoby, Z.-G. Wang, D. Weitz, E. Vlieg, J. DeYoreo, D. Kashchiev, G. Nicolis, J. Lutsko, and many others. While their contributions are reflected above, the responsibility for all omissions of facts and slips of logic is entirely mine. Our work on the nucleation of crystals and the physical chemistry of protein solutions has been funded by NHLBI-NIH, NSF, the Petroleum Research Fund of ACS, NASA, and The Welch Foundation. I thank NSF (Grant MCB 0843726) and The Normal Heckerman Advanced Research Program (Grant 003652-0078-2009) for financial support for the work on the paper.

References

1. J. J. De Yoreo, A. K. Burnham, and P. K. Whitman, Developing KH_2PO_4 and KD_2PO_4 crystals for the world's most power laser. *Int. Mater. Rev.*, **47**, 113–152 (2002).
2. P. Reichert, et al., Metal-interferon-alpha crystals, US Patent 5,441,734 (1995).
3. J. Brange, *Galenics of Insulin*, Springer, 1987.
4. M. L. Long, et al., Protein crystal growth in microgravity review of large scale temperature induction method. *J. Crystal Growth*, **168**, 233–243 (1996).
5. S. Matsuda, et al., New crystal form of recombinant murine interferon-β. *J. Biol. Chem.*, **264**, 13381–13382 (1989).
6. S. Peseta, J. A. Langer, K. C. Zoon, and C. E. Samuel, in *Annual Review of Biochemistry*, Vol. **56**, C. C. Richardson, P. D. Boyer, I. B. Dawid, and A. Meister, eds., Annual Reviews, 1989, pp. 727–778.
7. S. Charache, C. L. Conley, D. F. Waugh, R. J. Ugoretz, and J. R. Spurrell, Pathogenesis of hemolytic anemia in homozygous hemoglobin C disease. *J. Clin. Invest.*, **46**, 1795–1811 (1967).

8. R. E. Hirsch, C. Raventos-Suarez, J. A. Olson, and R. L. Nagel, Ligand state of intraerythrocytic circulating HbC crystals in homozygote CC patients. *Blood*, **66**, 775–777 (1985).
9. W. A. Eaton and J. Hofrichter, in *Advances in Protein Chemistry*, Vol. 40, C. B. Anfinsen, J. T. Edsal, F. M. Richards, and D.S. Eisenberg, eds., Academic Press, 1990, pp. 63–279.
10. P. Vekilov, Sickle-cell haemoglobin polymerisation: is it the primary pathogenic event of sickle-cell anaemia? *Br. J. Haematol.*, **139**, 173–184 (2007).
11. C. R. Berland, et al., Solid–liquid phase boundaries of lens protein solutions. *Proc. Natl. Acad. Sci. USA*, **89**, 1214–1218 (1992).
12. N. Asherie, et al., Enhanced crystallization of the Cys18 to Ser mutant of bovine gammaB crystallin. *J. Mol. Biol.*, **314**, 663–669 (2001).
13. G. Dodson, and D. Steiner, The role of assembly in insulin's biosynthesis. *Curr. Opin. Struct. Biol.*, **8**, 189–194 (1998).
14. A. McPherson, *Introduction to Macromolecular Crystallography*, Wiley, 2009.
15. H. M. Berman, et al., The Protein Data Bank. *Nucl. Acids Res.*, **28**, 235–242 (2000).
16. J. Prausnitz, and L. Foose, Three frontiers in the thermodynamics of protein solutions. *Pure Appl. Chem.*, **79**, 1435–1444 (2007).
17. K. S. Schmitz, *Dynamic Light Scattering by Macromolecules*, Academic Press, 1990.
18. D. N. Petsev, et al., Temperature-independent solubility and interactions between apoferritin monomers and dimers in solution. *J. Crystal Growth*, **232**, 21–29 (2001).
19. N. Takeno, *Open File Report of GSJ*, **49**, 26–143 (1988).
20. J. W. Gibbs, On the equilibrium of heterogeneous substances. *Trans. Connect. Acad. Sci.*, **3**, 108–248 (1876).
21. J. W. Gibbs, On the equilibrium of heterogeneous substances. *Trans. Connect. Acad. Sci.*, **16**, 343–524 (1878).
22. A. Milchev, Electrochemical phase formation on a foreign substrate—basic theoretical concepts and some experimental results. *Contemp. Phys.*, **32**, 321–332 (1991).
23. A. Milchev, *Elecrocrystallization: Fundamentals of Nucleation and Growth*, Kluwer, 2002.
24. E. D. Shchukin, A. V. Pertsov, and E. A. Amelina, *Colloid Chemistry*, Moscow University Press, 1982.
25. D. Kashchiev, *Nucleation. Basic Theory with Applications*, Butterworth, Heinemann, 2000.
26. B. Mutaftschiev, in *Handbook of Crystal Growth*, Vol. I, D. T. J. Hurle, ed., Elsevier, 1993, pp. 189–247.
27. M. Volmer, *Kinetik der Phasenbildung*, Steinkopff, 1939.
28. P. G. Vekilov, L. A. Monaco, B. R. Thomas, V. Stojanoff, and F. Rosenberger, Repartitioning of NaCl and protein impurities in lysozyme crystallization. *Acta Crystallogr. Sect. D*, **52**, 785–798 (1996).
29. J. Lothe, and G. M. Pound, On the statistical mechanics of nucleation theory. *J. Chem. Phys.*, **45**, 630–634 (1966).
30. M. L. Broide, C. R. Berland, J. Pande, O. O. Ogun, and G. B. Benedek, Binary liquid phase separation of lens proteins solutions. *Proc. Natl. Acad. Sci. USA*, **88**, 5660–5664 (1991).
31. M. Muschol, and F. Rosenberger, Liquid–liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation/crystallization. *J. Chem. Phys.*, **107**, 1953–1962 (1997)
32. D. N. Petsev, X. Wu, O. Galkin, and P. G. Vekilov, Thermodynamic functions of concentrated protein solutions from phase equilibria. *J. Phys. Chem. B*, **107**, 3921–3926 (2003).

33. P. E. Bonnett, K. J. Carpenter, S. Dawson, and R. J. Davey, Solution crystallization via a submerged liquid–liquid phase boundary: oiling out. *Chem. Commun.*, 698–699 (2003).
34. Y. B. Zel'dovich, Theory of new phase formation: cavitation. *Acta Physicochim. URSS*, **18**, 1–22 (1943).
35. D. Kashchiev, in *Science and Technology of Crystal Growth*, J. P. van der Eerdenand O. S. L. Bruinsma, eds., Kluwer Academic Publishers, 1995, pp. 53–56.
36. P. R. ten Wolde, and D. Frenkel, *Enhancement of protein crystal nucleation by critical density fluctuations. Science*, **277**. 1975–1978 (1997)
37. V. Talanquer, and D. W. Oxtoby, Crystal nucleation in the presence of a metastable critical point. *J. Chem. Phys.*, **109**, 223–227 (1998).
38. K. G. Soga, J. M. Melrose, and R. C. Ball, Metastable states and the kinetics of colloid phase separation. *J. Chem. Phys.*, **110**, 2280–2288 (1999).
39. J. A. Thomson, P. Schurtenberger, G. M. Thurston, and G. B. Benedek, Binary liquid phase separation and critical phenomena in a protein water solution. *Proc. Natl. Acad. Sci. USA*, **84**, 7079–7083 (1987).
40. N. Asherie, A. Lomakin, and G. B. Benedek, Phase diagram of colloidal solutions. *Phys. Rev. Lett.*, **77**, 4832–4835 (1996).
41. D. W. Oxtoby, *Crystals in a flash. Nature*, **420**, 277–278 (2002).
42. P. G. Vekilov, Dense liquid precursor for the nucleation of ordered solid phases from solution. *Crystal Growth Design*, **4**, 671–685 (2004).
43. O. Galkin, and P. G. Vekilov, Control of protein crystal nucleation around the metastable liquid–liquid phase boundary. *Proc. Natl. Acad. Sci. USA*, **97**, 6277–6281 (2000).
44. V. J. Anderson, and H. N. W Lekkerkerker, *Insights into phase transition kinetics from colloid science. Nature*, **416**, 811–815 (2002).
45. O. Galkin, and P. G. Vekilov, Direct determination of the nucleation rate of protein crystals. *J. Phys. Chem.*, **103**, 10965–10971 (1999).
46. P. G. Vekilov, and O. Galkin, On the methods of determination of homogeneous nucleation rates of protein crystals. *Colloids Surfaces A*, **215**, 125–130 (2003).
47. V. Bhamidi, S. Varanasi, and C. A. Schall, Measurement and modeling of protein crystal nucleation kinetics. *Crystal Growth Design*, **2**, 395–400 (2002).
48. N. M. Dixit, A. M. Kulkarni, and C. F. Zukoski, Comparison of experimental estimates and model predictions of protein crystal nucleation rates. *Colloids Surfaces A*, **190**, 47–60 (2001).
49. N. M. Dixit, and C. F. Zukoski, Crystal nucleation rates for particles experiencing short-range attractions: applications to proteins. *J. Colloid Interface Sci.*, **228**, 359–371 (2000).
50. O. Galkin, and P. G. Vekilov, Are nucleation kinetics of protein crystals similar to those of liquid droplets? *J. Am. Chem. Soc.*, **122**, 156–163 (2000).
51. W. Pan, A. B. Kolomeisky, and P. G. Vekilov, Nucleation of ordered solid phases of protein via a disordered high-density state: phenomenological approach. *J. Chem. Phys.*, **122**, 174905 (2005).
52. D. Kashchiev, On the relation between nucleation work, nucleus size, and nucleation rate. *J. Chem. Phys.*, **76**, 5098–5102 (1982).
53. D. W. Oxtoby, and D. Kashchiev, A general relation between the nucleation work and the size of the nucleus in multicomponent nucleation. *J. Chem. Phys.*, **100**, 7665–7671 (1994).
54. I. J. Ford, Nucleation theorems, the statistical mechanics of molecular clusters, and a revision of classical nucleation theory. *Phys. Rev. E*, **56**, 5615–5629 (1997).
55. J. N. Schmeltzer, Comments on the nucleation theorem. *J. Colloid Interface Sci.*, **242**, 354–372 (2001).

56. J. D. van der Waals, *Nobel Lectures, Physics* 1901–1921, Elsevier, 1910.
57. P. G. Debenedetti, *Metastable Liquids*, Princeton University Press, 1996.
58. J. W. Cahn, and J. E. Hilliard, Free energy of a nonuniform system. I. Interfacial free energy. *J. Chem. Phys.*, **28**, 258–267 (1958).
59. J. S. Langer, in *Fluctuations and Instabilities in Phase Transitions*, T. Riske, ed., Plenum Press, 1975, pp. 19–42.
60. L. Filobelo, *Kinetics of Phase Transition in Protein Solutions on Microscopic and Mesoscopic Length Scales*, Ph.D. thesis, University of Houston, 2005.
61. E. Cacioppo, and M. L. Pusey, The solubility of the tetragonal form of hen egg white lysozyme from pH 4.0 to 5.4. *J. Crystal Growth*, **114**, 286–292 (1991).
62. S. B. Howard, P. J. Twigg, J. K. Baird, and E. J. Meehan, The solubility of hen-egg-white lysozyme. *J. Crystal Growth*, **90**, 94–104 (1988).
63. L. F. Filobelo, O. Galkin, and P. G. Vekilov, Spinodal for the solution-to-crystal phase transformation. *J. Chem. Phys.*, **123**, 014904 (2005).
64. P. G. Vekilov, in *Perspectives on Inorganic, Organic and Biological Crystal Growth: From Fundamentals to Applications: AIP Conference Proceedings*, Vol. 916, *AIP Conference Series*, M. Skowronski, J. J. DeYoreo, and C. A. Wang, eds., AIP, 2007, pp. 235–267.
65. A. E. S Van Driessche, et al., Comparison of different experimental techniques for the measurement of crystal growth kinetics. *Crystal Growth Design*, **8**, 4316–4323 (2008).
66. A. J. Malkin, Y. G. Kuznetsov, and A. McPherson, *In situ* atomic force microscopy studies of surface morphology, growth kinetics, defect structure and dissolution in macromolecular crystallization. *J. Crystal Growth*, **196**, 471–488 (1999).
67. P. G. Vekilov, B. R. Thomas, and F. Rosenberger, Effects of convective solute and impurity transport on protein crystal growth. *J. Phys. Chem. B*, **102**, 5208–5216 (1998).
68. M. Shah, O. Galkin, and P. G. Vekilov, Smooth transition from metastability to instability in phase separating protein solutions. *J. Chem. Phys.*, **121**, 7505–7512 (2004).
69. L. K. Steinrauf, Preliminary X-ray data for some crystaline forms of β-lactoglobulin and hen egg-white lysozyme. *Acta Crystallogr.* **12**, 77–78 (1959).
70. A. J. Malkin, and A. McPherson, Light scattering investigations of protein and virus crystal growth: ferritin, apoferritin and satellite tobacco mosaic virus. *J. Crystal Growth*, **128**, 1232–1235 (1993).
71. A. J. Malkin, and A. McPherson, Light scattering investigation of the nucleation processes and kinetics of crystallization in macromolecular systems. *Acta Crystallogr. Sect. D*, **50**, 385–395 (1994).
72. D. Vivares, E. Kaler, and A. Lenhoff, Quantitative imaging by confocal scanning fluorescence microscopy of protein crystallization via liquid-liquid phase separation. *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 819–825 (2005).
73. O. Galkin, K. Chen, R. L. Nagel, R. E. Hirsch, and P. G. Vekilov, Liquid–liquid separation in solutions of normal and sickle cell hemoglobin. *Proc. Natl. Acad. Sci. USA*, **99**, 8479–8483 (2002).
74. W. Pan, O. Galkin, L. Filobelo, R. L. Nagel, and P. G. Vekilov, Metastable mesoscopic clusters in solutions of sickle cell hemoglobin. *Biophys. J.*, **92**, 267–277 (2007).
75. W. Pan, P. G. Vekilov, and V. Lubchenko, The origin of anomalous mesoscopic phases in protein solutions. *J. Phys. Chem. B*, **114**, 7620–7630 (2010).
76. O. Gliko, et al., Metastable liquid clusters in super- and undersaturated protein solutions. *J. Phys. Chem. B*, **111**, 3106–3114 (2007).

77. W. Pan, et al., Viscoelasticity in homogeneous protein solutions. *Phys. Rev. Lett.*, **102**, 058101 (2009).

78. O. Gliko, et al., A metastable prerequisite for the growth of lumazine synthase crystals. *J. Am. Chem. Soc.*, **127**, 3433–3438 (2005).

79. E. E. Uzgiris, and D. C. Golibersuch, Excess scattered-light intensity fluctuations from hemoglobin. *Phys. Rev. Lett.*, **32**, 37–40 (1974).

80. F. Sciortino, S. Mossa, E. Zaccarelli, and P. Tartaglia, Equilibrium cluster phases and low-density arrested disordered states: the role of short-range attraction and long-range repulsion. *Phys. Rev. Lett.*, **93**, 055701 (2004).

81. J. Groenewold, and W. K. Kegel, Anomalously large equilibrium clusters of colloids. *J. Phys. Chem. B*, **105**, 11702–11709 (2001).

82. Y. Liu, W. Chen,-R., and S. Chen,-H. Cluster formation in two-Yukawa fluids. *J. Chem. Phys.*, **122**, 044507 (2005).

83. S. Mossa, F. Sciortino, P. Tartaglia, and E. Zaccarelli, *Ground-state clusters for short-range attractive and long-range repulsive potentials. Langmuir*, **20**, 10756–10763 (2004).

84. A. Stradner, et al., *Equilibrium cluster formation in concentrated protein solutions and colloids. Nature*, **432**, 492–495 (2004).

85. S. B. Hutchens, and Z. Wang,-G. Metastable cluster intermediates in the condensation of charged macromolecule solutions. *J. Chem. Phys.*, **127**, 084912 (2007).

86. D. Kashchiev, Thermodynamically consistent description of the work to form a nucleus of any size. *J. Chem. Phys.*, **118**, 1837–1851 (2003)

87. W. J. Fredericks, M. C. Hammonds, S. B. Howard, and F. Rosenberger, Density, thermal expansivity, viscosity and refractive index of lysozyme solutions at crystal growth concentrations. *J. Crystal Growth*, **141**, 183–192 (1994).

88. A. A. Chernov, *Modern Crystallography III, Crystal Growth*, Springer, 1984.

89. M. A. Lauffer, Entropy-driven processes in biology. *Mol. Biol. Biophys.*, **20**, 1–264 (1975).

90. A. Hill, Entropy production as the selection rule between different growth morphologies. *Nature*, **348**, 426–428 (1990).

91. A. G. Walton, in *Nucleation*, A. C. Zettlemoyer, ed., Marcel Dekker, 1969, pp. 225–307.

92. A. Fersht, *Structure and Mechanism in Protein Science*, W.H. Freeman, 1999.

93. M. D. Hollingsworth, Crystal engineering: from structure to function. *Science*, **295**, 2410–2413 (2002).

94. J. S. Lee, Y. Lee,-J., E. L. Tae, Y. S. Park, and K. B. Yoon, Synthesis of zeolite as ordered multicrystal arrays. *Science*, **301**, 818–821 (2003).

95. A. McPherson, and P. Shlichta, The use of heterogeneous and epitaxial nucleants to promote the growth of protein crystals. *J. Crystal Growth*, **90**, 47–50 (1988).

96. S. A. Hemming, et al., The mechanism of protein crystal growth from lipid layers. *J. Mol. Biol.*, **246**, 308–316 (1995).

97. B. Garetz, J. Matic, and A. Myerson, Polarization switching of crystal structure in the nonphotochemical light-induced nucleation of supersaturated aqueous glycine solutions. *Phys. Rev. Lett.*, **89**, 175501 (2002).

98. Y. G. Kuznetsov, A. J. Malkin, and A. McPherson, The liquid protein phase in crystallization: a case study intact immunoglobins. *J. Crystal Growth*, **232**, 30–39 (2001).

99. O. Galkin, R. L. Nagel, and P. G. Vekilov, The kinetics of nucleation and growth of sickle cell hemoglobin fibers. *J. Mol. Biol.*, **365**, 425–439 (2007).

100. O. Galkin, et al., Two-step mechanism of homogeneous nucleation of sickle cell hemoglobin polymers. *Biophys. J.*, **93**, 902–913 (2007).

101. A. Lomakin, D. S. Chung, G. B. Benedek, D. A. Kirschner, and D. B. Teplow, On the nucleation and growth of amyloid b-protein fibrils: deterction of nuclei and quatification of rate constants. *Proc. Natl. Acad. Sci. USA*, **93**, 1125–1129 (1996).

102. R. Krishnan, and S. L. Lindquist, *Structural insights into a yeast prion illuminate nucleation and strain diversity. Nature*, **435**, 765–772 (2005).

103. J. E. Aber, S. Arnold, and B. A. Garetz, Strong dc electric field applied to supersaturated aqueous glycine solution induces nucleation of the polymorph. *Phys. Rev. Lett.*, **94**, 145503 (2005).

104. M. E. Leunissen, et al., *Ionic colloidal crystals of oppositely charged particles. Nature*, **437**, 235–240 (2005).

105. J. R. Savage, and A. D. Dinsmore, Experimental evidence for two-step nucleation in colloidal crystallization. *Phys. Rev. Lett.*, **102**, 198302 (2009).

106. T. H. Zhang, and X. Y. Liu, Multistep crystal nucleation: a kinetic study based on colloidal crystallization. *J. Phys. Chem. B*, **111**, 14001–14005 (2007).

107. L. B. Gower, Biomimetic model systems for investigating the amorphous precursor pathway and its role in biomineralization. *Chem. Rev.*, **108**, 4551–4627 (2008).

108. E. M. Pouget, et al., *The initial stages of template-controlled $CaCO_3$ formation revealed by cryo-TEM. Science*, **323**, 1455–1458 (2009).

109. D. Gebauer, A. Volkel, and H. Colfen, *Stable prenucleation calcium carbonate clusters. Science*, **322**, 1819–1822 (2008)

110. J. F. Wang, M. Muller, and Z. G. Wang, Nucleation in A/B/AB blends: interplay between microphase assembly and macrophase separation. *J. Chem. Phys.*, **130**, (2009).

111. R. P. Sear, Nucleation via an unstable intermediate phase. *J. Chem. Phys.*, **131**, 074702 (2009).

112. J. F. Lutsko, and G. Nicolis, Theoretical evidence for a dense fluid precursor to crystallization. *Phys. Rev. Lett.*, **96**, 046102 (2006).

EXPERIMENTAL STUDIES OF TWO-STEP NUCLEATION DURING TWO-DIMENSIONAL CRYSTALLIZATION OF COLLOIDAL PARTICLES WITH SHORT-RANGE ATTRACTION

JOHN R. SAVAGE,[1,2] LIQUAN PEI,[1] AND ANTHONY D. DINSMORE[1]

[1]*Department of Physics, University of Massachusetts, Amherst, MA 01003, USA*
[2]*Department of Physics, Cornell University, Ithaca, NY 14853, USA*

CONTENTS

I. INTRODUCTION

The process by which atoms, molecules, or particles form crystals after a temperature quench is important in many natural situations and applications but still the subject of active investigations [2–4]. The rate of crystallization is limited by

Advances in Chemical Physics, Volume 151: Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanoscale Materials, First Edition. Edited by Gregoire Nicolis and Dominique Maes.

the early stages, known as nucleation, in which thermal fluctuations lead to small clusters of the new crystalline phase. The classical nucleation theory assumes that small clusters form with the same structure as the new phase but that the interfacial energy leads to a rate-limiting free energy barrier and a minimum size for stable clusters [2–5]. This model was originally envisioned for nucleation of liquid droplets in a gas phase, but is often extended to the case of solids nucleating in a gas or liquid phase. By this mechanism, a small cluster having the structure of the new crystalline phase forms spontaneously. While this cluster is small, its free energy exceeds that of the initial phase because of an energy penalty at the cluster boundary (a surface energy term). While most of these clusters minimize their free energy by breaking up, random collisions make a few clusters large enough that the change in bulk energy exceeds the change in surface energy, and the clusters can thereafter minimize their free energy by growing. In this model, the rate at which stable clusters are formed is then related to the probability that random fluctuations lead to a cluster of sufficient size: the rate is proportional to the exponential of the free energy barrier. While it is plausible that the classical single-step pathway might be the most rapid one, it is now known that this is not always the case. In one alternative scenario, nucleation may occur by the sample's passing through multiple states in order of decreasing free energy [3, 6–8]. This proposed mechanism is most often referred to as Ostwald's rule of stages, which proposes that a sequence of crystalline phases should appear on supercooling a fluid [6]. A beautiful example of Ostwald's rule is shown by a recent high-resolution electron microscopy study of crystallization of $LiFePO_4$ [9, 10]. In other cases, however, nucleation of a solid phase might involve noncrystalline phases, with fluid or liquid crystalline symmetry. Moreover, how the nature of the interactions affects the nucleation dynamics is not well established.

Crystallization of globular proteins is a critical step in finding protein structure and recent work on this topic has advanced our understanding of crystallization from a supercooled liquid or gas. In 1997, ten Wolde and Frenkel described Monte Carlo computer simulations, in which globular proteins in solution were modeled as spheres having a hard-core repulsion and a short-range attraction [7, 11]. They showed that under these circumstances nucleation does not always proceed by a classical single-step process. Instead, in a narrow range of temperatures and concentrations, nucleation can proceed in two distinct steps. In the simulations, droplets of a liquid phase, enriched in protein, appeared first. At a crossover size of approximately 200 particles, these liquid droplets then became crystalline. Until this point, the clusters would still most often break up because their size was below the critical value at which the free energy reaches a peak, but thermal fluctuations led some clusters to grow beyond the critical size. This two-step nucleation process is illustrated in Fig. 1. More recent computer simulations [4, 11–14] and theoretical models including density functional theory, liquid structure theory, and kinetic models [15–20] have similarly predicted that two-step nucleation should

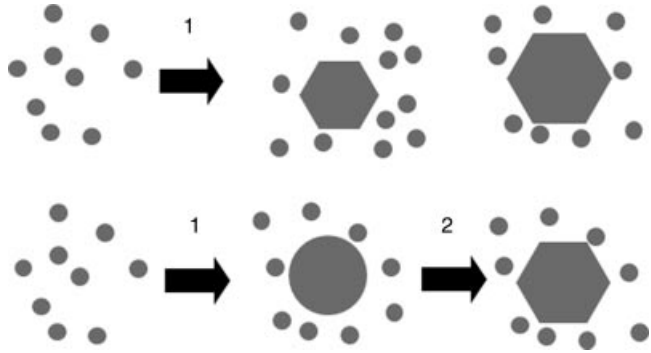


Figure 1. Schematic illustrations of (top) the classical nucleation model and (bottom) two-step nucleation.

occur. Indeed, density functional theory [19] and computer simulations [14, 21] more recently predicted that two-step nucleation should also occur in the Lennard-Jones system. Hence, the phenomenon of two-step nucleation during crystallization might be quite common.

For predicting crystallization rates, the two-step nucleation process is crucial because it significantly reduces the free energy barrier, which causes an exponentially large enhancement of the rate of crystallization [11, 19, 20]. In some systems, the nucleation pathway might also affect the long-time behavior if the sample becomes trapped in one of the nucleation steps [22–25].

Despite the fact that phase transitions occur out of equilibrium and involve relatively small clusters of particles, properties of the *bulk equilibrium* phase diagram are nonetheless informative [6, 8]. A key distinction between atomic systems and globular proteins is the range of the attraction. Many atomic systems (such as Ar) can be reasonably modeled by a Lennard-Jones or similar interaction, in which particles strongly repel one another near contact and attract one another at longer distances over a range that is comparable to the particle diameter. By contrast, proteins (and the colloidal particles in our experiments) have a range of attraction that is much smaller than the particle size. This distinction has a major effect on the equilibrium phase diagram in two- and three-dimensional systems. An atomic (Lennard-Jones-like) system exhibits the three familiar phases: gas (g), liquid (l), and crystalline solid (s) (Fig. 2a). This diagram has the familiar g-l critical point (•), triple temperature (—), and regions of two-phase coexistence. For short-range attractive systems (Fig. 2b), there are only two phases: fluid and crystal. Curiously, no g-l coexistence exists in equilibrium because the g-l binodal is superseded by the g-s binodal. In Fig. 2, the metastable g-l region is indicated by the shading. Microscopically, the liquid's metastability may be explained by the fact that when a crystal melts into a liquid, many short-range contacts are lost and the lost bond

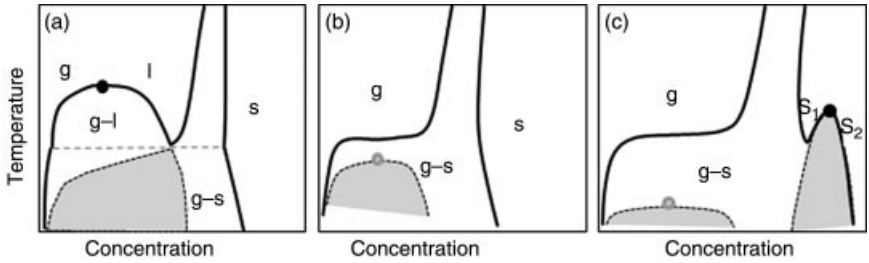


Figure 2. Schematic phase diagrams for systems with isotropic hard-core repulsion and longer range attraction. (**a**) When the range of attraction is comparable to the particle size (as in a Lennard-Jones system), the phase diagram has gas (g), liquid (l), and crystalline solid (s) phases, with regions of coexistence. The gray region denotes where g–l coexistence is metastable because the liquid is superseded by the crystal. (**b**) For a short-range attraction, only g and s remain and g–l coexistence and the critical point become metastable (gray). (c) For an even shorter range of attraction (<10% of particle diameter), a second solid phase (s_1) appears along with a new s_1–s_2 critical point. Adapted from Ref. 1.

energy overwhelms the gain in entropy. If the attraction is of long range, then many of these contacts persist in the liquid phase. Experiments with colloidal particles with variable interactions support these predicted phase diagram topologies [26].

When a sample is cooled from gas to the region of g-s coexistence, nucleation may be affected by the metastable g-l critical point and the metastable g-l spinodal region. Near this region of parameter space, density fluctuations cost little energy (or minimize energy, in the spinodal region), and these fluctuations occur immediately on supercooling. Crystallization is now more likely to occur in the dense regions, where the probability that random fluctuations lead to a large crystalline cluster is much larger. In terms of the macroscopic phase diagram, the nucleating clusters pass through the metastable liquid phase on their way to the solid phase. In the Lennard-Jones-type system, the g-l critical point lies well outside the l-s region, but a supercooled sample still forms metastable liquid clusters before crystallizing [14, 19, 21].

When the attraction has a still shorter range (<10% of the particle diameter), a new crystalline phase (s_1) is predicted to appear with the same symmetry as the first solid but at a higher concentration. This new phase gives rise to another critical point, as shown in Fig. 2c. To our knowledge, this isostructural solid–solid coexistence (s_1–s_2) has not been experimentally observed, possibly because it requires a high concentration of monodisperse particles. Nonetheless, the existence of this second solid phase can affect the dynamics of freezing and melting, even in samples with concentrations well below the s_1–s_2 coexistence. Cacciuto et al. found in simulations that spheres with very short-range attractions nucleate first in the metastable s_1 phase, and then continuously evolve into the equilibrium s_2 phase [27].

Experimental studies of crystallization of globular proteins provide evidence of the role of the metastable liquid phase. In protein solutions, crystallites nucleate within metastable liquid-phase droplets (glucose isomerase [28]) or evolve from metastable clusters (sickle cell hemoglobin [29] or lysozyme [30]). Moreover, the nucleation rate density in lysozyme solutions reaches a maximum near the metastable gas–liquid (g-l) boundary [31]; this is consistent with the prediction that two-step nucleation lowers the free energy barrier. These experiments suggest that two-step nucleation is occurring over a wide region of phase space (wider, in fact, than initially predicted [11, 15] but consistent with more recent predictions of Lutsko and Nicolis [19]). Owing to the small particle size, however, these experiments could not probe the evolving structure at the scale of individual particles and clusters.

Experiments with colloidal particles offer the opportunity to probe phase transitions with single-particle resolution and with tunable interparticle interactions. This approach has been exploited for several years [3] to study melting with attractive or repulsive interactions [24, 32–36] and freezing of nearly hard spheres [37, 38]. Studies of colloidal particles in 2D that attract one another by electric field-induced flow revealed two-step nucleation of crystallites within dense liquid droplets. The interaction, however, was of long range and possibly dependent on cluster size, and the fluid flow might alter the dynamics [39]. Closer to the topic of this chapter, Hobbie probed freezing in colloids confined to two dimensions with short-range attraction and reported a two-step freezing process with amorphous clusters appearing first and crystalline clusters later (when the area fraction was $<30\%$) [24]. At larger area fractions, the metastable amorphous phase was arrested. The dynamics of the formation of these liquid-like clusters and their size distribution were analyzed by Smoluchowski dynamics. Here, however, we evaluate cluster size distributions from a thermodynamic view and focus on the role of the liquid clusters in the crystallization process.

In this chapter, we describe our experimental studies of two-dimensional crystallization of colloidal particles with a tunable short-range attractive interaction. These experiments consist of "supercooling" samples via small temperature changes and monitoring the size and evolving symmetry of individual clusters. Figure 3 shows an image of a gas-phase sample following a temperature quench. Initially, the sample is a supercooled, metastable gas of particles (Fig. 3a), but after a few minutes the sample has formed large numbers of hexagonally ordered crystallites (Fig. 3b). These samples are polycrystalline and a slow process of coarsening continues for at least several days. Here, we focus on the nucleation process—the rate-limiting events that occur between Fig. 3a and b. In previous publications [35, 40], we reported evidence of two-step melting and nucleation in this system; here, we provide a more detailed investigation. From a statistical analysis of the size and crystallinity of clusters, we measure their free energies and identify which samples lie within the metastable g-l region. We show that the

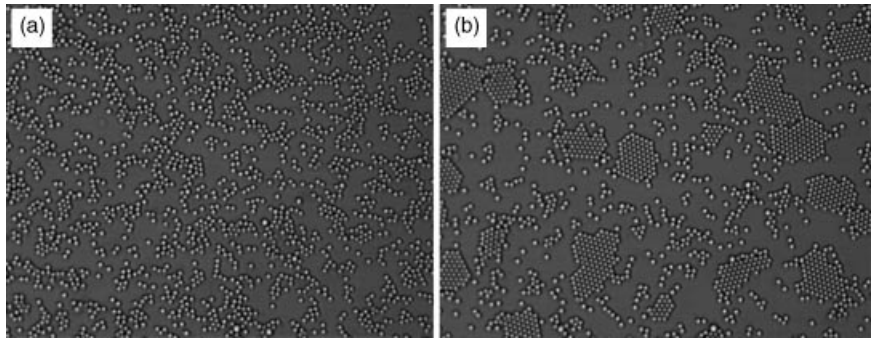


Figure 3. Image of a sample following a temperature "quench" into the gas crystal coexistence region. (**a**) Soon after the temperature change, the sample remains a metastable gas. (**b**) After 30 min, the sample has formed several crystallites. These crystallites subsequently undergo a slow process of growth. The white features are polystyrene spheres with a diameter of 1.4 μm and an area fraction of 24%. This chapter addresses the process by which crystallites such as these nucleate.

nucleation process consists of a single step (as in classical nucleation theory) at the lowest and highest particle concentrations probed here (17% and 55%). Over a range of intermediate densities, however, we observe two-step nucleation. Our data indicate that the two-step process has a lower energy barrier because of the large energy of the crystal–gas interface. We also find that samples cooled to within the metastable g-l coexistence region have a much higher nucleation rate density than do samples outside this region. Taken together, the results are consistent with the previously predicted two-step, gas–liquid–solid process and indicate the region of the phase diagram in which this phenomenon occurs. Remarkably, the process can be usefully described in terms of these *macroscopic* phases even when the clusters are as small as 10 particles. Finally, in the concluding section, we describe how these results may be relevant to crystallization in three dimensions and in systems with different forms of interaction.

II. EXPERIMENTAL METHODS: SAMPLE PREPARATION AND IMAGING

We used video microscopy to track charge stabilized spheres dispersed in Millipore-filtered water. The experiments reported here used either of two types of particles, polystyrene or silica. The polystyrene spheres were purchased from Interfacial Dynamics (Portland, OR, USA). The average sphere radius, R, was 0.7 μm with a polydispersity of 3.5% of the mean, according to the manufacturer. The silica particles were purchased from Polysciences, Inc. (Warrington, PA, USA) and

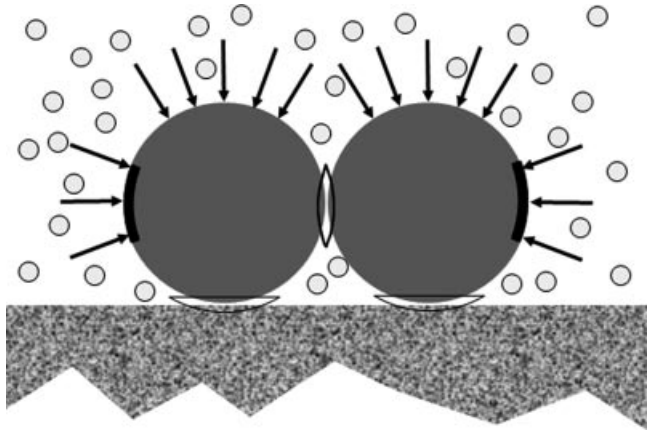


Figure 4. Illustration of the mechanism of depletion attraction. The small particles (micelles, in the present case) apply an osmotic pressure on the surfaces of the larger spheres. In the regions where the two spheres nearly touch one another or the flat surface, the centers of mass of the micelles are excluded. (These exclusion regions are shown by the lens-shaped features). Because of these exclusion regions, the osmotic pressure of the micelles is imbalanced and a net pressure is applied as shown by the arrows. The result is that spheres are attracted to the flat surface and to other nearby spheres. The range of the interaction is set by the diameter of the micelles and the strength is set by their diameter and concentration.

had a diameter that was nominally 1.5 μm with a polydispersity of approximately 3% of the mean.

To induce a pairwise attraction between particles and to confine the particles to a single layer, we take advantage of the "depletion" or "macromolecular crowding" effect (Fig. 4) [35, 41–43]. To all of our samples, we added the nonionic surfactant hexaethylene glycol monododecyl ether ($C_{12}E_6$) at a concentration of approximately 4.4×10^{-2} M. This surfactant forms micelles when the concentration exceeds the critical micelle concentration, which is approximately 8.7×10^{-5} M at 25°C. The magnitude of the depletion interaction can be described in terms of the change of Helmholtz free energy (ΔF_d) when two spheres move from far apart to contact: $\Delta F_d/(k_B T) \approx -2\pi R R_m(T)^2 c_m(T)$, where $R_m(T)$ is the micelle radius and $c_m(T)$ is the concentration of micelles (not the critical micelle concentration) [35, 40]. For the samples and temperatures described here, $R_m(T) = 9$–17 nm (measured with dynamic light scattering). The concentration of added surfactant is not precisely known, however, because $C_{12}E_6$ is hygroscopic and the concentration of the stock surfactant solution gradually changed over time. Even if the magnitude of ΔF_d is not known, the key point for these experiments is that it can be varied by means of temperature. For nonionic surfactants such as $C_{12}E_6$, the critical micelle concentration decreases with increasing T [44], so more of the

surfactant molecules form micelles and c_m increases. In addition, $R_m(T)$ increases with T [45]. By a combination of these effects, heating enhances the depletion interaction and causes crystallization. It might seem surprising that samples are homogeneous fluids at low T and have crystals at higher T; the reason is simply that heating scarcely changes the thermal energy $k_B T$, but it significantly strengthens ΔF_d. For the remainder of this chapter, we will refer to a temperature change that induces crystallization as a "supercooling" into the gas–solid coexistence region, even though it actually corresponds to a rising T.

Samples were mixed, and then injected into a glass chamber for viewing on the microscope. A glass coverslip (#1.5) was used for the bottom surface for high-resolution imaging and a standard glass slide was used for the top. Careful cleaning of all glass pieces was important for reliable results. Our cleaning procedure was to soak the glass for several hours in a mixture of concentrated sulfuric acid with "nochromix" powder added (Sigma-Aldrich, cat. no. 328693). (*Note*: this procedure uses corrosive liquid and is hazardous. A vented chemical hood, protective gear, and proper procedures should be used.) Afterward, the glass was rinsed thoroughly in Millipore-filtered water and dried in an oven. Epoxy was used to seal the coverslip to the slide, with a gap of approximately 250 μm and two holes remaining in the epoxy seal. After the epoxy had set, the sample was injected through one of the holes, after which the holes were sealed with epoxy. These samples typically lasted for several days without noticeable evaporation or leakage.

In order to form crystals, we first formed a two-dimensional, weakly interacting gas of spheres on the coverslip. Because the mass density of the polystyrene particles (1.055 g cm^{-3}) is slightly larger than water, the particles slowly sediment toward the coverslip. The depletion attraction between each sphere and the flat surface is approximately $2 \times \Delta F_d$ [35, 43], so the particles become bound to the coverslip but are still mobile in two dimensions. Once all the particles were bound to the coverslip, a second increase of T initiated nucleation and growth. In samples composed of polystyrene spheres, we were unable to achieve a particle area fraction in excess of 34% without having many particles in a second layer. To obtain higher area fractions, we used silica spheres, taking advantage of the greater mass density of silica, which leads to a single layer with few second-layer particles even when $\eta > 55\%$.

Images were acquired at 30 s^{-1} using a Zeiss Axiovert 200 inverted optical microscope with a 63× Plan-Neofluar oil coupled objective (numerical aperture = 1.4), a CCD video camera, an S-VHS recorder, and a frame grabber attached to a computer. The microscope was placed on a mechanically isolated table to suppress vibrations. A box made of 2 in. thick rigid foam insulation surrounded the microscope. Samples were heated with an objective heater (Zeiss Tempcontrol 37). Data were acquired after enough time had passed for T to stabilize with temperature fluctuations approximately 0.1°C.

III. ANALYSIS OF IMAGES: PARTICLE COORDINATES, AREA FRACTION, CLUSTER SIZES, AND CRYSTALLINE ORDER

The positions of the particle centroids were measured by computer analysis of the images using routines written in IDL (ITT Visual Information Solutions, Boulder CO, USA) [46]. When imaging with visible light at high magnification, it is possible to measure the centroids of spheres with an uncertainty that is far smaller than the Rayleigh resolution limit [46–48], even as small as 1 nm [49]. For the present experiments, the resolution of the particle centroid finding algorithm was tested by measuring the particles' mean square displacement as a function of time, and then extrapolating to zero time delay. The random error in the centroid positions thus obtained was typically $(\langle \Delta x^2 \rangle + \langle \Delta y^2 \rangle)^{1/2} = 47$ nm.

For each sample studied, we define the area fraction of particles, η, as $N a_p/A$ (expressed as a percentage), where N is the counted number of particles, a_p is the area of each particle (πR^2), and A is the total area of the imaged region. In our samples, η ranged between 17% and 55%.

To quantify the nucleation and growth of individual clusters, we use two parameters: one that measures only size and the other that measures only crystalline order. A cluster is defined as the set of N particles that are contiguously bonded, where a "bond" exists between any two particles whose center-to-center distance is within a cutoff length [35]. This definition is particularly useful for studying phase transitions because it does not discriminate between highly ordered "crystalline" particles and less ordered or "liquid-like" particles. For the polystyrene spheres (used for most of the samples discussed here), we chose a cutoff length of 1.53 μm. This length is slightly longer than the range of the depletion attraction, $2(R + R_S) \sim 1.43$ μm; the longer cutoff length accounts for the random measurement error that would otherwise cause us to miss bonds. While alterations of the value of the cutoff length slightly change the values of N, the trends in the data reported here are not altered. The sample with area fraction 55% was composed of silica spheres that are slightly larger than the polystyrene spheres, so we adjusted the cutoff length to 1.63 μm. The number of particles in a cluster is defined as N. Crystallinity of clusters was quantified using the standard bond orientational order parameter, ψ_6. Following the established convention, we define for each particle (labeled by the index j) $|\psi_6(j)|^2 \equiv (1/Z_j)|\Sigma_k \exp[6i\theta_{jk}]|^2$, where Z_j is the number of bonds with the jth particle. The sum extends over all bonded neighbors k, and θ_{jk} is the angle between the j-k bond and the x-axis (see Fig. 5). By computing the modulus, the result for $|\psi_6|^2$ is independent of the choice of reference axis. With this definition, any particle whose bonds have perfect triangular symmetry has $|\psi_6|^2 = 1.0$, even if $Z < 6$. By contrast, particles with random bond angles have $|\psi_6|^2$ approaching 0. Particles that have $Z = 1$ are considered noncrystalline and $\psi_6 \equiv 0$ for these cases.

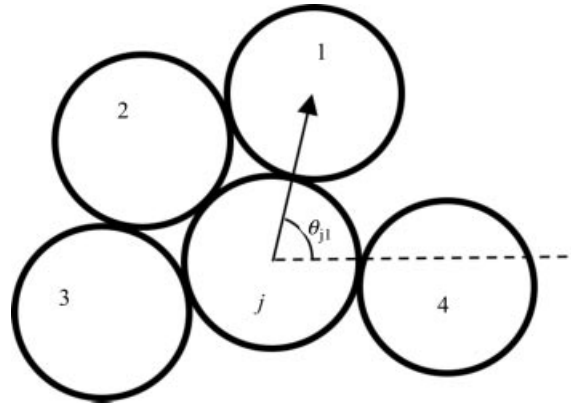


Figure 5. Illustration of the definition of the bond orientational order parameter $\psi_6(j)$. Here, the jth particle has four neighbors.

To characterize the crystallinity of clusters, we compute the average $|\psi_6|^2$ for all particles within the cluster; we label this parameter $\langle|\psi_6|^2\rangle$. $\langle|\psi_6|^2\rangle_{\mathrm{av}}(N)$ is defined as $\langle|\psi_6|^2\rangle$ averaged over all clusters of size N in the sample.

IV. RESULTS

In this section, we describe the results of our experiments with colloidal particles. We begin in Section IV.A with a description of the samples in the first minutes following supercooling, during which time we find many liquid-like clusters throughout the samples. We compare these clusters with the subpopulation of clusters that eventually become stable crystallites. In Section IV.B, we describe measurements of the free energy as a function of cluster size. This analysis indicates which samples lie within the metastable g-l coexistence region and provides estimates of the line tension of the g-l and g-s interfaces. In Section IV.C, we describe time-averaged images that indicate liquid cluster formation in low-concentration samples and gas bubble formation in high-concentration samples. In Section IV.D, we describe the nucleation rate density and its dependence on the sample's location in the phase diagram.

A. Single-Step and Two-Step Nucleation of Crystallites

For an extended period of time following supercooling, our samples contained a large number of disordered clusters with a wide range of sizes. Figure 3a shows an image of the sample with $\eta = 24\%$. In this case, more than a third of the particles were incorporated into clusters of $N > 1$. These clusters varied widely in size, from $N = 2$ to $N > 50$. The particles in these clusters were mobile and rapidly exchanged neighbors with a characteristic timescale of several seconds. In most samples, the

clusters themselves typically survived for a period of a few tens of seconds before spontaneously breaking up into much smaller clusters and monomers. (In the two most strongly supercooled samples, with $\eta = 28.5\%$ and 34%, the clusters tended to grow rather than breaking up, as will be discussed later.) The characteristic lifetime of the clusters is considerably longer than the time required for an individual particle to diffuse a distance equal to its own diameter ($R^2/D \sim 1.6$ s), so the clusters do not arise simply from random collisions but are instead held together by attractive interactions. In samples with η less than approximately 34%, these clusters were distinct entities with much greater concentration than the background.

Inspection of the image in Fig. 3a shows that these early time clusters were disordered, lacking any discernible crystalline order. The cluster-averaged bond orientational order parameter, $\langle|\psi_6|^2\rangle_{\rm av}$, is also relatively low, approximately 0.4 (Fig. 6). By contrast, the crystallites that appear later have $\langle|\psi_6|^2\rangle_{\rm av} \approx 0.8$, as will be shown later. The data of Fig. 6 were acquired during the relatively early times after supercooling (typically within 200 s after the temperature had stabilized),

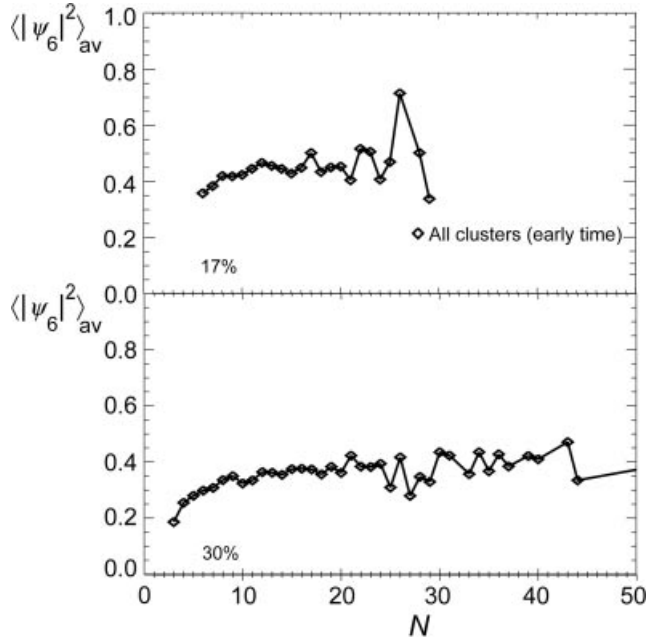


Figure 6. Plot of the average crystallinity of clusters that appear after "supercooling" and prior to the formation of stable crystallites. The vertical axis shows the bond orientational order parameter $|\psi_6|^2$ averaged over all particles in a cluster. The plots show data for two samples with area fractions $\eta = 17\%$ and 30%.

prior to the formation of crystallites. We verified that the distribution of cluster sizes and the values of $\langle|\psi_6|^2\rangle_{\text{av}}$ did not vary significantly over this time, so the cluster statistics represent a steady-state property.

Because of the fluid-like mobility within the clusters and the lack of bond orientational order, we interpret these clusters as droplets of the liquid phase. In most samples, these droplets were unstable and rapidly vaporized; in other samples, these droplets were metastable, eventually changing to crystalline clusters. While the cluster size would appear to be too small to justify rigorously a continuum treatment, we find nonetheless that the stability of these clusters corresponds to the predicted properties of the bulk liquid phase.

We now turn to the relatively small number of clusters that continued to grow in size and eventually became stable crystallites. These stable clusters were readily identified by the observation that once they formed, they persisted for the remainder of the experiment (approximately 0.5 h or longer). From our images, we identify these clusters once they become large and stable, and then trace them backward in time to monitor their evolution. This analysis allows us to focus exclusively on the pathway that leads to crystals.

Figure 7 summarizes the distinction between the metastable clusters and those clusters that become large and stable. In this plot, the diamonds (-◇-) show the clusters observed prior to the appearance of stable clusters (the same data as in Fig. 6, but now for five different samples). The solid lines show the clusters that eventually grew to a size of 100 or greater and then continued growing. At large N (or late time), all these curves approach $\langle|\psi_6|^2\rangle_{\text{av}} > 0.8$ and the clusters appear ordered to the eye (e.g., see inset of Fig. 7d). Moreover, the particles in these clusters maintain their neighbors for timescales that exceed the observation time (hours). The finite size makes the question of their long-term translational order ill-defined (and in any case true long-range translational order does not exist in two dimensions [50] even though there is a first-order phase transition [51]). We refer to these clusters as crystallites.

A key result of these studies is that the pathway to forming large, stable crystallites varies significantly depending on both η and depth of supercooling. In the 17% sample, even small "successful" clusters had $\langle|\psi_6|^2\rangle_{\text{av}} > 0.7$, even at small size ($N < 10$). That is, these clusters were already structurally distinct from the amorphous clusters of the background and already had the order characteristic of the late-stage crystallites. This behavior is consistent with the classical nucleation theory, in which nucleation proceeds in a single step (see Fig. 1a).

At higher concentration, however, we find a different nucleation pathway. In the sample with $\eta = 30\%$, the growing clusters passed through two distinct stages of growth (Fig. 7d). In the first stage ($N < 20$), the clusters were liquid-like and indistinguishable from the metastable clusters in terms of crystalline order and the mobility of individual particles. As N approached 30 particles, there was a distinct increase in the value of $\langle|\psi_6|^2\rangle_{\text{av}}$, with clusters becoming ordered when

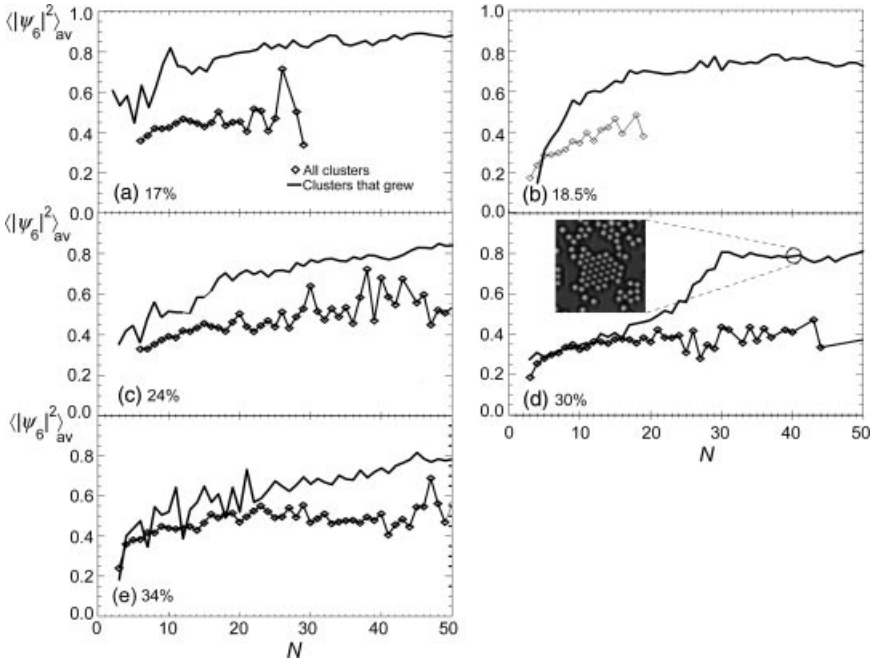


Figure 7. Plot of the average crystallinity of clusters that appeared in five supercooled samples with different particle area fraction, η. As in Fig. 6, diamonds show the data for all the amorphous clusters that appeared prior to stable clusters. The solid curves show the data for those clusters that grew to large size and persisted for the remainder of the experiment.

$N > 30$ ($\langle|\psi_6|^2\rangle_{\rm av} \sim 0.8$). This is a clear example of two-step nucleation. Behavior intermediate between classical nucleation-like and two steps were exhibited by a range of samples.

Might this two-step behavior be explained by wetting of the crystalline phase by a liquid? By this wetting mechanism, the interior of a cluster would be crystalline and the perimeter would contain disordered, liquid-like particles. At small N, these perimeter particles would comprise the majority of the cluster and cause it to appear disordered. As clusters grow, the fraction of highly ordered particles would increase and $\langle|\psi_6|^2\rangle_{\rm av}$ increase, even as the liquid-like layer remained. To investigate the possibility of wetting, we identified particles as being "perimeter-like" if they had fewer than six bonds ($Z < 6$). We then separately tracked the ordering of these particles. Following our earlier convention in melting experiments [35], we find all particles j in each cluster with $Z < 6$. We then calculate the local orientational order parameter $C_6(j) \equiv \langle|(1/Z_{\rm j})\Sigma_{\rm k}\psi_6(j)\psi_6{}^*(k)|^2\rangle$, where $\Sigma_{\rm k}$ refers to a sum over all neighbors (k) of j regardless of Z_k. We then calculate the cluster-averaged value

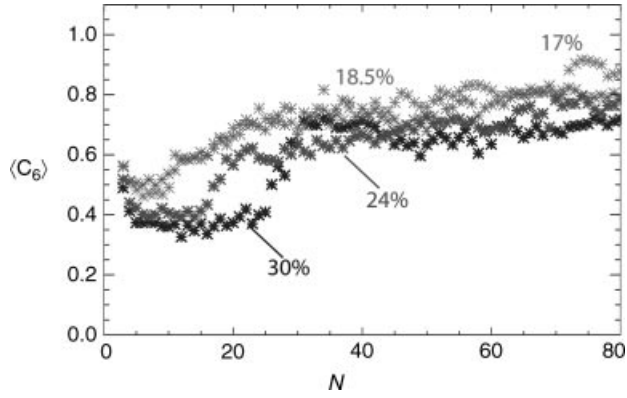


Figure 8. Plot of the measured correlation of each particle's bond orientational order with its neighbors' order, $\langle C_6 \rangle$. This plot shows data only for particles that have fewer than six bonds ("perimeter"-like particles). In the samples with $\eta = 24$ and 30%, these particles suddenly become ordered when N reaches 20–30.

of C_6, which we term $\langle C_6 \rangle$. Figure 8 shows $\langle C_6 \rangle$ as a function of cluster size N in successfully growing clusters, considering only those particles with $Z < 6$. The sudden onset of order for the 30% sample occurs at $N \sim 25$–30, showing that ordering occurs for even "perimeter"-like and not just core particles having $Z = 6$. The disordered-to-order transition occurs throughout the cluster. A similar result is seen in the sample with $\eta = 24\%$ at a crossover size of $N \sim 20$. The samples with $\eta = 17\%$ and 18.5% had little or no crossover in the $\langle C_6 \rangle$ parameter.

B. The Free Energy as a Function of Cluster Size

During the interval of time between supercooling and the formation of stable crystallites (often more than 100 s), the liquid-like clusters were highly dynamic, but in most cases the statistical distribution of cluster sizes and cluster-averaged $|\psi_6|^2$ were constant. Figure 9 shows a plot of the mean cluster size $\langle N \rangle$ as a function of time in several different samples. Most of the samples shown here (17%, 18.5%, 24%, and 30%) maintained an approximately constant $\langle N \rangle$ for the first 200 s. We conclude that these systems underwent thermal fluctuations (clustering and breakup) about a steady state. This implies that the distribution of cluster sizes follows the Boltzmann distribution: the number of clusters of size N should be $n_N \propto \exp(-\Delta F(N)/k_B T)$, where ΔF is the difference in Helmholtz free energy between N particles in the gas and a *liquid* cluster of size N. Since we can readily measure n_N, we invert this formula to find $\Delta F(N)/k_B T = -\ln(n_N/n_1)$, where n_1 is the number of monomers (unbonded particles). (Normalization by n_1 conveniently sets $\Delta F = 0$ for a monomer.)

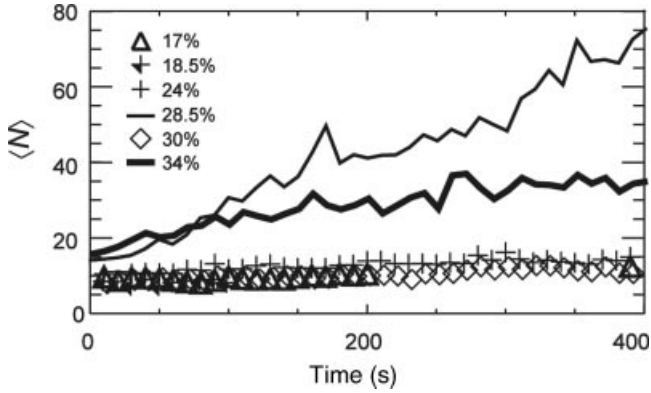


Figure 9. Plot of the mean cluster size $\langle N \rangle$ versus time. The time $t = 0$ correspond to the start of the recording, where temperature reached a constant value.

Two samples—those with $\eta = 28.5\%$ and 34%—showed a continuous growth of $\langle N \rangle$ and a slight change in the cluster size distribution over time. In these cases, the clusters were not strictly growing and shrinking in steady state, so the assumption that $\Delta F(N)/k_B T = -\ln(n_N/n_1)$ is not exact. Nonetheless, we assume that because the overall n_N changes slowly, the energy landscape also changes only slightly during the first 100 s of video.

Figure 10 shows a plot of $-\ln(n_N/n_1)$ versus N for six samples. These data correspond to the early stages of nucleation (typically the first 100 s), during which time there were no stable, growing clusters. Hence, these data describe the first step of nucleation: the formation of (metastable) liquid clusters. As might be expected, these clusters have higher free energy than the gas ($\Delta F > 0$), so the great majority of them eventually evaporate.

To interpret these results, we assume that these liquid clusters can be described by a simplified classical nucleation theory: $\Delta F = (\Delta\mu_{l-g})N + ({\eta_c}^{1/2}\pi\Gamma_{l-g})N^{1/2}$. Here, $\Delta\mu_{l-g}$ is the chemical potential of the liquid minus that of the gas. The second term represents the interfacial energy of the liquid cluster; Γ_{l-g} is the line tension in units of $k_B T$ per sphere diameter and η_c is the area fraction of particles in the cluster. When $\Delta\mu_{l-g} > 0$, this function monotonically rises, so liquid clusters would never be stable. For $\Delta\mu_{l-g} < 0$, the function would first rise and then decrease when N exceeds a critical size; this is the typical scenario in classical nucleation. In Fig. 10a, we show the result of a best fit to the simplified classical nucleation model, which shows good agreement with the data. For clarity, we show the fit only for $\eta = 30\%$; the best-fit parameters are $\Delta\mu_{l-g} = 0.04 \pm 0.02 k_B T$. To obtain estimates of the numerical values of the line tension, we simplify the analysis by assuming that clusters were circular and $\eta_c = (3/\pi)^2 = 0.91$. It should be noted, however,

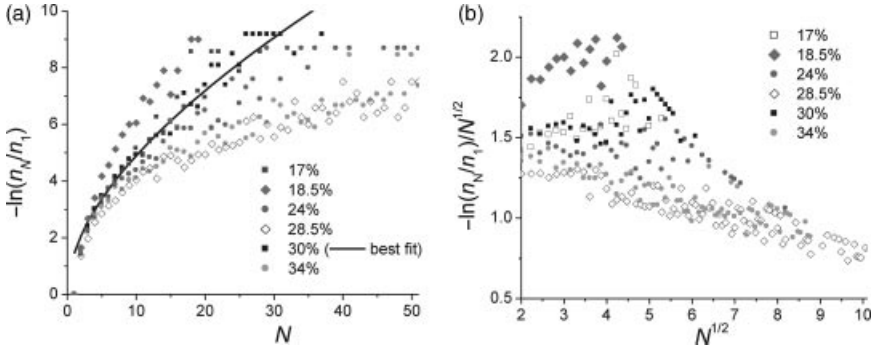


Figure 10. (**a**) Plot of the free energy versus cluster size N for samples with different area fractions η and T. Since these samples maintained a steady-state distribution of cluster sizes, we assumed that the fluctuations followed the Boltzmann distribution and hence the vertical axis is the free energy of a cluster, in units of k_BT. These clusters are disordered, and therefore indicate the free energy of the liquid phase. (**b**) Rescaled plot, showing $F/\sqrt{N}$ versus $\sqrt{N}$. By classical nucleation theory, $F/\sqrt{N} = (\Delta\mu_{l-g})\sqrt{N}$ plus a constant related to line tension. The slope of the data indicates whether the sample lies within the metastable g-l coexistence region.

that these clusters are not always circular, and there is good evidence that classical nucleation in general would not lead to circular or spherical clusters [5] and the numerical value of Γ_{l-g} should therefore be interpreted with care. By this best-fit analysis, the 28.5% and 34% samples both had $\Gamma_{l-g} = 0.48\,k_BT$/diameter and $\Delta\mu_{l-g} = -0.06(1)k_BT$ for $t < 100$ s.

To extract the sign of $\Delta\mu_{l-g}$ more directly, we also plot $-\ln(n_N/n_1)/N^{1/2}$ versus $N^{1/2}$ (Fig. 10b). According to the classical model, the result should be a straight line with slope $\Delta\mu_{l-g}/k_BT$ and y-intercept ${\eta_c}^{1/2}\pi\Gamma$. The data are consistent with this scaling for N larger than approximately 4. This plot shows clearly that the samples with $\eta = 28.5\%$ and 34% both have $\Delta\mu_{l-g} < 0$, indicating that these particles lay inside the metastable g-l coexistence regions. A *sufficiently large* liquid cluster would spontaneously grow, but the global stability of the liquid phase is superseded by the *crystal* phase once N is sufficiently large.

By contrast, the samples with $\eta = 17\%$, 18.5%, and 30% lay outside the g-l region (i.e., $\Delta\mu_{l-g} > 0$). Most of these liquid clusters evaporated eventually as thermal fluctuations continually made new clusters. The sample with $\eta = 24\%$ lay very close to the metastable g-l binodal ($\Delta\mu_{l-g} \approx 0$).

C. Locating Samples on the Phase Diagram

Further insight into the role of the metastable liquid phase is obtained by a coarse-grained view of the supercooled samples. Samples with low η tended to have dense clusters in a dilute background, but the highest η sample showed a structure more accurately described as dilute "voids," akin to gas bubbles. To identify these

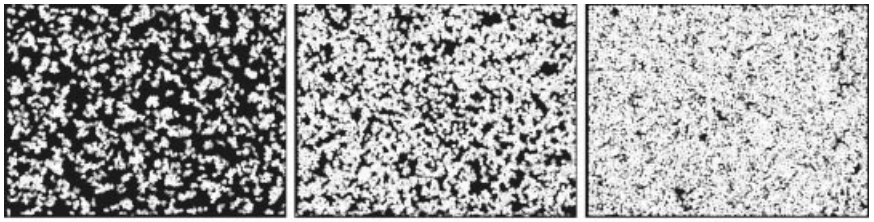


Figure 11. Time-averaged images of three samples prior to formation of stable crystallites. The image each averages 10 s of video as described in the text; the white regions are particles and the black are voids. (**a**) The cluster-in-gas structure is clearly visible at area fraction $\eta = 17\%$. (**b**) At $\eta = 28.5\%$, the clusters and voids have nearly equal area fraction. (**c**) At the relatively high fraction, $\eta = 55\%$, there are voids in a concentrated background. The clusters at low η and the voids at high η persist for tens of seconds.

structures more clearly, we first created images in which the particles were represented by white circles on a black background (using the particle positions known from image analysis) and then calculated the time average of these reconstructed images from $t = 0$ to $t = 10$ s. In Fig. 11, we show these images for samples with $\eta = 17\%$, 28.5%, and 55%.

At the lowest concentration, we find dense liquid clusters (white) surrounded by a dilute, gas-like background (black). As described previously, these clusters persisted for roughly several tens of seconds before breaking up into smaller clusters and monomers. In the language of continuum thermodynamics, the clusters are unstable liquid droplets appearing spontaneously in the gas phase. This behavior would be expected for a sample that lies outside the g-l binodal on the low-concentration side.

In the strongly supercooled 28.5% sample, we find comparable areas of "cluster" and "void" regions. In terms of continuum thermodynamics, this structure resembles the bicontinuous gas–liquid structure characteristic of spinodal decomposition. (We point out, however, that we do not currently have sufficient statistics to establish this point without ambiguity.) This type of morphology is consistent with our finding that $\Delta\mu_{\mathrm{l-g}} < 0$, so the sample was supercooled into the metastable g-l region. These void regions persisted in time and coarsened as crystallites appeared.

At the highest concentration studied here ($\eta = 55\%$), the situation is reversed: here, we find small discrete voids in a concentrated liquid background. The time lapse movies show that these voids persisted for roughly 20 s before vanishing. Again, this behavior can be described in the language of continuum thermodynamics as gas "bubbles" that are unstable with respect to the surrounding liquid phase. The bubbles form, but rapidly vanish. This behavior would be expected for a system that lies just outside the g-l binodal on the high-concentration side. (Alternatively, it is possible that this sample lay within the metastable g-l region

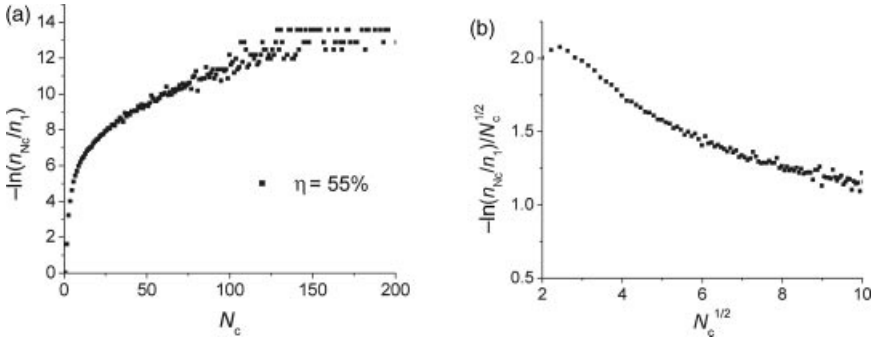


Figure 12. (**a**) Plot of the free energy versus size of crystalline clusters, N_c, for the sample with $\eta = 55\%$. (**b**) The same data, plotted as $\Delta F/\sqrt{N_c}$ versus $\sqrt{N_c}$. The negative slope indicates that the chemical potential is lower in the crystalline phase than in the initial fluid phase.

but that the gas bubbles (voids) never reached the critical size before the sample crystallized.) In this sample, crystallization can be viewed as classical nucleation of the crystal directly from the liquid phase.

Following the classical nucleation model for the sample with $\eta = 55\%$, we measured the distribution of *crystalline* clusters. We began by calculating $|\psi_6|^2$ for each particle, and then defined as "crystalline" those particles with $|\psi_6|^2 \geq 0.90$. (This choice of cutoff value is justified by the fact that the distribution of $|\psi_6|^2$ values in crystallizing samples typically has a sharp peak near 1; this peak is distinct from the broad background distribution.) We then looked for contiguous clusters of these "crystalline" particles and defined the size of each cluster as N_c. We defined n_{Nc} as the number of clusters of size N_c and then found the free energy of crystalline clusters from $-\ln(n_{Nc})$ (Fig. 12). The rescaled plot shows an approximate consistency with the classical nucleation model, $\Delta F = (\Delta\mu_{s-l})N_c + (\eta_c{}^{1/2}\pi\Gamma_{s-l})N_c{}^{1/2}$ with $\Delta\mu_{s-l} < 0$. From the extrapolated y-intercept, we estimate the line tension as $\Gamma_{s-l} \approx 0.8\, k_B T$/diameter. Interestingly, this line tension is approximately 50% larger than Γ_{l-g} (Fig. 10). We will return to this point in Section V.

D. Measured Nucleation Rates

Because the rate of nucleation is expected to depend exponentially on the free energy barrier, it is usually the case that the rate should vary significantly. To measure the nucleation rate density, we begin at the end of each of our videos and count the number of stable, persistent clusters. We then play the movie backward and count the number of such clusters at earlier times. Our images show a sample area of $1.37 \times 10^4\ \mu m^2$, so the data may be converted to density.

In Fig. 13, we show the number of persistent clusters (which are all crystalline) as a function of time for seven samples with η ranging from 17% to 55%. The nucleation rate density varied widely. The most rapid nucleation occurred in the

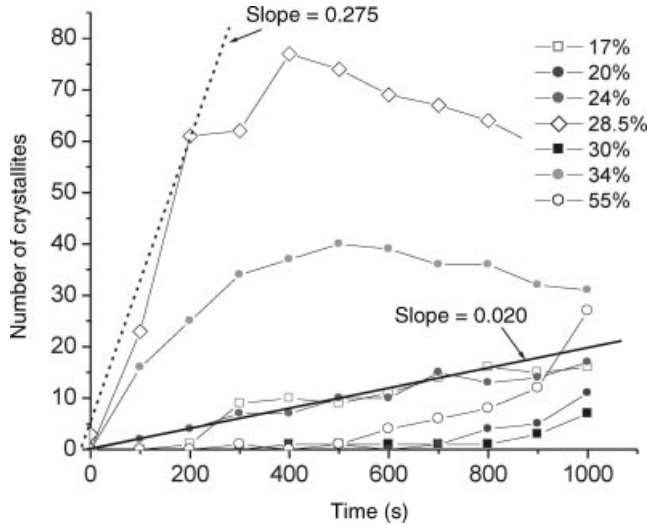


Figure 13. Plot of the number of observed stable (growing) crystallites in the full field of view as a function of time.

28.5% sample, in which 60 stable crystallites appeared within the first 3 min. The 34% sample contained 25 crystallites over the same time and area. These are the two samples for which $\Delta\mu_{\mathrm{l-g}} < 0$, so it seems clear that the metastable gas–liquid region enhanced the nucleation rate. By contrast, the other samples contained at most four stable crystallites after 3 min, and some samples showed no stable crystallites for more than 5 min. The rate is not simply related to η, since the most concentrated sample was one of the slowest ones. Rather, it appears that proximity to the g-l coexistence region is the key factor, and the nucleation rate was enhanced by approximately an order of magnitude by entering the metastable g-l binodal. This result is consistent with simulations [11, 12] and density functional theory [15, 19], which predict large enhancement of the crystallization rate near the metastable g-l critical point. They are also broadly consistent with the experiments with proteins [31], which were done systematically as a function of T. In our experiments, however, it is clear that two-step nucleation occurs even in samples that lay outside the metastable g-l coexistence region (e.g., $\eta = 30\%$).

For an overview of the relative roles of cluster size and crystalline symmetry of the "successful" clusters during their nucleation, we collected statistics of both the total cluster size N and the size of the *crystalline* part of the cluster, N_{c}. (As above, N_{c} is defined as the set of particles for which $|\psi_6|^2 \geq 0.9$). For this analysis, we measured a two-dimensional histogram, collecting the number of clusters ($n_{N,N\mathrm{c}}$) with given N and N_{c}. The values N and N_{c} were measured for all successfully nucleating clusters in 10 s intervals. We assume that the system undergoes fluctuations about a stationary free energy function so that the logarithm of $n_{N,N\mathrm{c}}$ gives

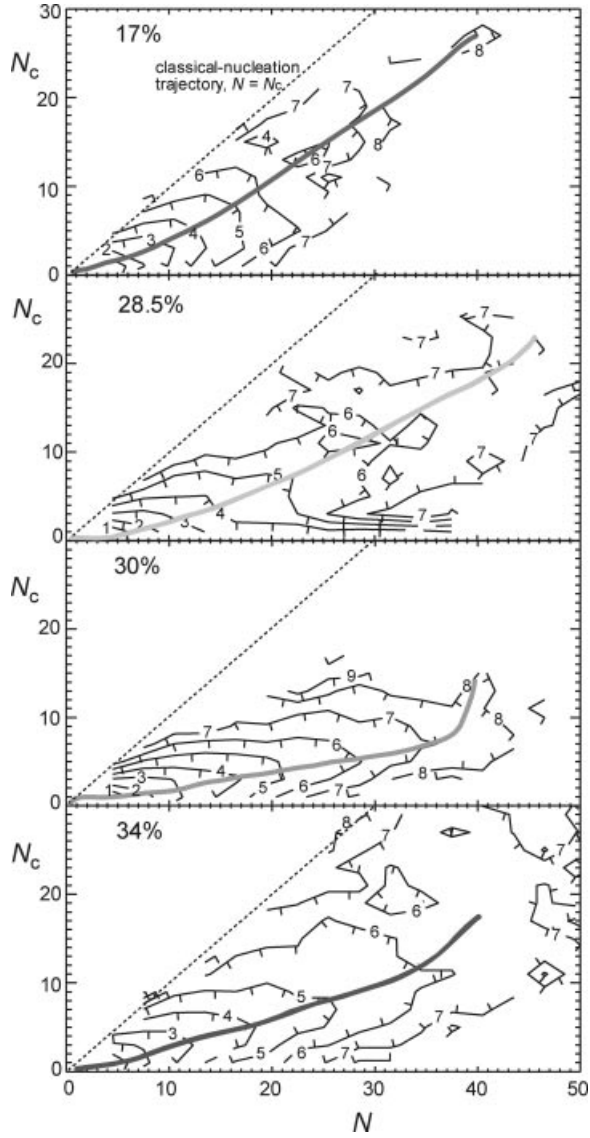


Figure 14. Plots of the free energy, $\Delta F/k_B T$, of successfully nucleating clusters are shown as a function of N and N_c for four samples. The black curves show contours of equal ΔF and the thicker, gray curves are drawn by eye to illustrate the most probable trajectories. The classical nucleation model would have $N = N_c$, which is indicated by the dashed line.

the (free energy)/$k_B T$. (A similar kind of analysis was reported by ten Wolde and Frenkel for short-range interactions [11] and by Moroni et al. to study nucleation in a Lennard-Jones liquid [52].)

Figure 14 shows the free energy plots measured from our data. The data are noisy because of limited statistics (especially near the edges of the plotted data),

but some important trends are nonetheless apparent. To identify what would plausibly be the most likely nucleation pathway, we have drawn curves that follow the gradient in the ΔF. For the sample with $\eta = 17\%$, this pathway has a nearly steady slope that resembles the classical model: size increases along with order. In the sample $\eta = 30\%$, however, the clusters tend first to grow in size, and then to become ordered. A similar trend was reported for nucleation in short-range systems in three dimensions (by ten Wolde and Frenkel [11]). Those authors found that near the metastable g-l critical point, a precritical liquid cluster developed crystalline order at a crossover size of approximately 200 particles. In the present experiments, the crossover size has a maximum value of approximately 35 in the 30% sample. These sizes (200 in 3D and 35 in 2D) correspond to different spatial dimensions and different systems, and they need not be related. Nonetheless, it may be significant that $(200)^{2/3} \approx 34$, which implies that the radial length of a cluster is the same in these two systems. Whether there exists a universal value of the crossover cluster near the metastable g-l point and the question of which parameters quantitatively determine this size are interesting topics of future research.

The free energy contours also reveal the extent to which the classical nucleation theory would likely overestimate the nucleation rate density. In the 30% sample, for instance, the classical pathway and the measured pathway probably differ by a few $k_B T$ (we lacked sufficient statistics to measure ΔF along the line $N = N_c$). For the 34% sample, the difference is approximately $2.5 k_B T$ when $N = 30$, which would correspond to an enhancement of the nucleation rate by a factor of approximately $e^{2.5} = 12$.

V. SUMMARY AND CONCLUSION

By direct visualization of the clusters throughout the nucleation process, we have measured the evolution of crystalline order for the ensemble of clusters and separately for the "successful" clusters, which eventually grew to large size and survived for the duration of our experiments. In samples with intermediate area fractions (roughly, $20\% \leq \eta \leq 34\%$), nucleation proceeded in two steps: liquid first, then solid. From the distribution of liquid cluster sizes, we measured the cluster free energies. By comparing the first of these two nucleation steps to the classical model for nucleation of a liquid in gas, we determine which samples have $\Delta\mu_{g-l} < 0$ that must lie within the (metastable) g-l coexistence region. Combining these data with the observed mesoscopic-scale structure of the 17% and 55% samples (Fig. 11), we conjecture that the samples are arranged in a phase diagram as shown in Fig. 15a. The sample that had the largest size for crossover from liquid to solid had $\eta = 30\%$, which we assume lies near the metastable critical point. Computer simulations of two-dimensional systems with short-range attraction provide an estimate of the g-l critical area fraction: $\eta_c \approx 40\%$ in 2D [51]. Possibly the critical η_c in our experiments is lower than in the simulations, or the effect of the metastable critical

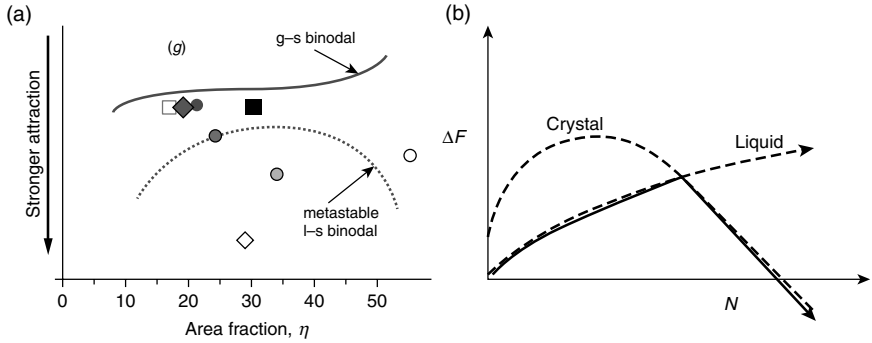


Figure 15. (**a**) Illustration of the conjectured phase diagram based on the reported behavior of the samples discussed here. (The plot symbols are the same as those used in Fig. 10.) (**b**) A schematic plot of the conjectured free energies for nucleating a crystalline cluster or a liquid droplet in a gas at moderate area fraction η. Because the crystal has larger line tension, it has a steeper curve at small size N. Hence, a liquid cluster is more readily formed even if it has positive $\Delta\mu$. The heavy solid curve shows the optimal trajectory.

point extends over a fairly broad region of phase space. Previous experiments with short-range depletion attraction in dilute samples ($\eta = 2\%$ in two dimensions [53] and 16% in 2D [54]) showed nucleation of crystals after a shallow quench, which appears consistent with our reported classical nucleation at 17%. These prior experiments, however, found that amorphous diffusion-limited structures appeared after a deeper quench. In addition, Hobbie also reported amorphous gel structures in samples with $\eta > 20\%$, with no crystals even after several hours [24]. These results are contrary to our results, since crystals eventually form in all of our samples. The difference in behavior might arise from difference in range of the potential (7% of the diameter for Ref. 24), salt concentration (stability against van der Waals aggregation), or polydispersity of the particle size.

To explain in general terms why samples undergo two-step nucleation, we return to the free energy measurements of Figs. 10 and 12. Monitoring the size of crystalline clusters led to a measurement of the gas crystal line tension that was approximately 50% greater than the gas–liquid line tension. This result might arise from the difference in symmetry in the g-s case. If we were to plot the free energy ΔF of a nucleating cluster, we would find something like the schematic shown in Fig. 15b, with a minimum free energy path that switches from the fluid branch to the crystal branch at a size that is larger than the critical size of a crystalline cluster. Qualitatively, this curve describes the behavior of the samples with intermediate area fraction (20–34%): in these samples once the cluster crystallized, it grew; crystalline clusters rarely broke apart. The absence of the two-step mechanism in the 17% sample might be explained by the relatively large and positive $\Delta\mu_{\mathrm{l-g}}$,

which would make fluid clusters too energetically costly. Further studies of the line tensions in various regions of the phase diagram are in progress.

In two-dimensional systems of spheres (or disks), the optimal short-range packing (hexagonal) is the same as the crystal packing [51]. This explains why crystallization occurred rapidly in our samples once the crystallites were energetically favorable. In three dimensions, the locally optimal packing is incommensurate with the crystal and crystallization might be slowed down [5, 55] so that the effect of two-step nucleation on the nucleation rate might be more dramatic. In this case, the fluid clusters might grow to even larger size before ordering, or alternatively they might become trapped in a disordered gel or glass [8, 53]. In three dimensions, multistep nucleation is also expected to be relevant among atoms and molecules with long-range interactions below the triple point [6, 14, 19, 21, 39, 56].

Acknowledgments

We thank Donald Blair, Alex Levine, Robert Guyer, and Daan Frenkel for valuable discussions. Funding for this work came from the National Science Foundation (DMR-0907195) and the Research Corporation through a Research Innovation Award.

References

1. D. Frenkel, Colloidal encounters: a matter of attraction, *Science*, **314**, 768 (2006).
2. A. Laaksonen, V. Talanquer, and D. W. Oxtoby, Nucleation: measurements, theory, and atmospheric applications, *Annu. Rev. Phys. Chem.*, **46**, 489 (1995).
3. V. J. Anderson and H. N. W. Lekkerkerker, Insights into phase transition kinetics from colloid science, *Nature*, **416**, 811 (2002).
4. R. P. Sear, Nucleation: theory and applications to protein solutions and colloidal suspensions, *J. Phys. Condens. Matter*, **19**, 033101 (2007).
5. F. Trudu, D. Donadio, and M. Parrinello, Freezing of a Lennard-Jones fluid: from nucleation to spinodal regime, *Phys. Rev. Lett.*, **97**, 105701 (2006).
6. J. W. Cahn, The metastable liquidus and its effect on the crystallization of glass, *J. Am. Ceram. Soc.*, **52**, 118 (1969).
7. P. R. ten Wolde and D. Frenkel, Homogeneous nucleation and the Ostwald step rule, *Phys. Chem. Chem. Phys.*, **1**, 2191 (1999).
8. F. Renth, W. C. K. Poon, and R. M. L. Evans, Phase transition kinetics in colloid-polymer mixtures at triple coexistence: kinetic maps from free-energy landscapes, *Phys. Rev. E*, **64**, 031402 (2001).
9. S. Y. Chung, Y. M. Kim, J. G. Kim, and Y. J. Kim, Multiphase transformation and Ostwald's rule of stages during crystallization of a metal phosphate, *Nat. Phys.*, **5**, 68 (2009).
10. S. J. L. Billinge, How do your crystals grow?, *Nat. Phys.*, **5**, 13 (2009).
11. P. R. ten Wolde and D. Frenkel, Enhancement of protein crystal nucleation by critical density fluctuations, *Science*, **277**, 1975 (1997).
12. K. G. Soga, J. R. Melrose, and R. C. Ball, Metastable states and the kinetics of colloid phase separation, *J. Chem. Phys.*, **110**, 2280 (1999).
13. A. Lomakin, N. Asherie, and G. B. Benedek, Liquid–solid transition in nuclei of protein crystals, *Proc. Natl. Acad. Sci. USA*, **100**, 10254 (2003).

14. J. A. van Meel, A. J. Page, R. P. Sear, and D. Frenkel, Two-step vapor-crystal nucleation close below triple point, *J. Chem. Phys.*, **129**, 204505 (2008).
15. V. Talanquer and D. W. Oxtoby, Crystal nucleation in the presence of a metastable critical point, *J. Chem. Phys.*, **109**, 223 (1998).
16. C. Haas and J. Drenth, The interface between a protein crystal and an aqueous solution and its effects on nucleation and crystal growth, *J. Phys. Chem. B*, **104**, 368 (2000).
17. A. Shiryayev and J. D. Gunton, Crystal nucleation for a model of globular proteins, *J. Chem. Phys.*, **120**, 8318 (2004).
18. D. Kashchiev, P. G. Vekilov, and A. B. Kolomeisky, Kinetics of two-step nucleation of crystals, *J. Chem. Phys.*, **122**, 244706 (2005).
19. J. F. Lutsko and G. Nicolis, Theoretical evidence for a dense fluid precursor to crystallization, *Phys. Rev. Lett.*, **96**, 046102 (2006).
20. N. M. Dixit and C. F. Zukoski, Crystal nucleation rates for particles experiencing short-range attractions: applications to proteins, *J.Colloid Interface Sci.*, **228**, 359 (2000).
21. B. Chen, H. Kim, S. J. Keasler, and R. B. Nellas, An aggregation-volume-bias Monte Carlo investigation on the condensation of a Lennard-Jones vapor below the triple point and crystal nucleation in cluster systems: an in-depth evaluation of the classical nucleation theory, *J. Phys. Chem. B*, **112**, 4067 (2008).
22. W. C. K. Poon, F. Renth, R. M. L. Evans, D. J. Fairhurst, M. E. Cates, and P. N. Pusey, Colloid-polymer mixtures at triple coexistence: kinetic maps from free-energy landscapes, *Phys. Rev. Lett.*, **83**, 1239 (1999).
23. P. J. Lu, E. Zaccarelli, F. Ciulla, A. B. Schofield, F. Sciortino, and D. A. Weitz, Gelation of particles with short-range attraction, *Nature*, **453**, 499 (2008).
24. E. K. Hobbie, Metastability and depletion-driven aggregation, *Phys. Rev. Lett.*, **81**, 3996 (1998).
25. M. G. Noro, N. Kern, and D. Frenkel, The role of long-range forces in the phase behavior of colloids and proteins, *Europhys. Lett.*, **48**, 332 (1999).
26. S. M. Ilett, A. Orrock, W. C. K. Poon, and P. N. Pusey, Phase behavior of a model colloid-polymer mixture, *Phys. Rev. E*, **51**, 1344 (1995).
27. A. Cacciuto, S. Auer, and D. Frenkel, Breakdown of classical nucleation theory near isostructural phase transitions, *Phys. Rev. Lett.*, **93**, 166105 (2004).
28. D. Vivares, E. W. Kaler, and A. M. Lenhoff, Quantitative imaging by confocal scanning fluorescence microscopy of protein crystallization via liquid–liquid phase separation, *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 819 (2005).
29. O. Galkin, W. Pan, L. Filobelo, R. E. Hirsch, R. L. Nagel, and P. G. Vekilov, Two-step mechanism of homogeneous nucleation of sickle cell hemoglobin polymers, *Biophys. J.*, **93**, 902 (2007).
30. A. M. Streets and S. R. Quake, Ostwald ripening of clusters during protein crystallization, *Phys. Rev. Lett.*, **104**, 178102 (2010).
31. O. Galkin and P. G. Vekilov, Control of protein crystal nucleation around the metastable liquid–liquid phase boundary, *Proc. Natl. Acad. Sci. USA*, **97**, 6277 (2000).
32. C. A. Murray and D. H. van Winkle, Experimental observation of two-stage melting in a classical two-dimensional screened Coulomb system, *Phys. Rev. Lett.*, **58**, 1200 (1987).
33. K. Zahn, R. Lenke, and G. Maret, Two-stage melting of paramagnetic colloidal crystals in two dimensions, *Phys. Rev. Lett.*, **82**, 2721 (1999).
34. A. M. Alsayed, M. F. Islam, J. Zhang, P. J. Collings, and A. G. Yodh, Premelting at defects within bulk colloidal crystals, *Science*, **309**, 1207 (2005).

35. J. R. Savage, D. W. Blair, A. J. Levine, R. A. Guyer, and A. D. Dinsmore, Imaging the sublimation dynamics of colloidal crystallites, *Science*, **314**, 795 (2006).

36. Y. Han, N. Y. Ha, A. M. Alsayed, and A. G. Yodh, Melting of two-dimensional tunable-diameter colloidal crystals, *Phys. Rev. E*, **77**, 041406 (2008).

37. U. Gasser, E. R. Weeks, A. Schofield, P. N. Pusey, and D. A. Weitz, Real-space imaging of nucleation and growth in colloidal crystallization, *Science*, **292**, 258 (2001).

38. Z. R. Wang, A. M. Alsayed, A. G. Yodh, and Y. L. Han, Two-dimensional freezing criteria for crystallizing colloidal monolayers, *J. Chem. Phys.*, **132**, 154501 (2010).

39. T. H. Zhang and X. Y. Liu, How does a transient amorphous precursor template crystallization, *J. Am. Chem. Soc.*, **129**, 13520 (2007).

40. J. R. Savage and A. D. Dinsmore, Experimental evidence for two-step nucleation in colloidal crystallization, *Phys. Rev. Lett.*, **102**, 198302 (2009).

41. A. Vrij, Polymers at interfaces and the interactions in colloidal dispersions, *Pure Appl. Chem.*, **48**, 471 (1976).

42. S. Asakura and F. Oosawa, Interaction between particles suspended in solutions of macromolecules, *J. Polymer Sci.*, **33**, 183 (1958).

43. P. D. Kaplan, J. L. Rouke, A. G. Yodh, and D. J. Pine, Entropically driven surface phase separation in binary colloidal mixtures, *Phys. Rev. Lett.*, **72**, 582 (1994).

44. L. J. Chen, S. Y. Lin, C. C. Huang, and E. M. Chen, Temperature dependence of critical micelle concentration of polyoxyethylenated non-ionic surfactants, *Colloid Surf. A*, **135**, 175 (1998).

45. B. Jönsson, B. Lindman, K. Holmberg, and B. Kronberg, *Surfactants and Polymers in Aqueous Solution*, Wiley, Chichester, 1998.

46. J. C. Crocker and D. G. Grier, Methods of digital video microscopy for colloidal studies, *J. Colloid Interface Sci.*, **179**, 298 (1996).

47. J. S. Olafsen, in: *Experimental and Computational Techniques in Soft Condensed Matter Physics*, J. S. Olafsen, ed., Cambridge University Press, Cambridge, 2010.

48. Y. X. Gao and M. L. Kilfoil, Accurate detection and complete tracking of large populations of features in three dimensions, *Opt. Express*, **17**, 4685 (2009).

49. J. Gelles, B. J. Schnapp, and M. P. Sheetz, Tracking kinesin-driven movements with nanometre-scale precision, *Nature*, **331**, 450 (1988).

50. P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics*, Cambridge University Press, 1995.

51. M. G. Noro and D. Frenkel, Phase behavior of a simple model for membrane proteins, *J. Chem. Phys.*, **114**, 2477 (2001).

52. D. Moroni, P. R. ten Wolde, and P. G. Bolhuis, Interplay between structure and size in a critical crystal nucleus, *Phys. Rev. Lett.*, **94**, 235703 (2005).

53. E. H. A. de Hoog, W. K. Kegel, A. van Blaaderen, and H. N. W. Lekkerkerker, Direct observation of crystallization and aggregation in a phase-separating colloid-polymer suspension, *Phys. Rev. E*, **6402**, 021407 (2001).

54. J. J. Cerda, T. Sintes, C. M. Sorensen, and A. Chakrabarti, Kinetics of phase transformations in depletion-driven colloids, *Phys. Rev. E*, **70**, 011405 (2004).

55. L. F. Filobelo, O. Galkin, and P. G. Vekilov, Spinodal for the solution-to-crystal phase transformation, *J. Chem. Phys.*, **123**, 014904 (2005).

56. X. B. Chen, A. C. S. Samia, Y. B. Lou, and C. Burda, Investigation of the crystallization process in 2 nm Cdse quantum dots, *J. Am. Chem. Soc.*, **127**, 4372 (2005).

ON THE ROLE OF METASTABLE INTERMEDIATE STATES IN THE HOMOGENEOUS NUCLEATION OF SOLIDS FROM SOLUTION

JAMES F. LUTSKO

Physics Department, Center for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, CP231, Blvd. du Triomphe, 1050 Brussels, Belgium

CONTENTS

Advances in Chemical Physics, Volume 151: Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanoscale Materials, First Edition. Edited by Gregoire Nicolis and Dominique Maes.

I. INTRODUCTION

The goal of this work is to describe, at a semimicroscopic level, the process of homogeneous nucleation of solids from solution. This is studied in the approximation that the effects of the solvent can be accounted for by an effective interaction between the solvate molecules so that at low concentrations, the molecules in solution are treated as a single-component gas with the crystalline phase being the corresponding solid phase in the effective single component system. In this picture, there is also the possibility of a dense liquid-like phase that simply corresponds to the liquid phase of the effective single-component system. Thus, the following work is intended to be applicable both to nucleation in single-component systems, such as simple fluids, and to nucleation of a solid phase from macromolecules in solution. For the sake of consistency, most of the discussion will be phrased in the language of nucleation of solids in a single component system. So, in the following, all statements concerning the "vapor phase" should be understood to be equally applicable to the "low-concentration" or "low-density" phase of molecules, such as proteins, in solution and statements concerning the "liquid phase" are equally applicable to the "high-concentration" or "high-density" phase of a solution.

In general, homogeneous nucleation involves at least two phases: the stable phase, S, and the metastable phase, M. Initially, the system consists of M in its homogeneous, bulk state. Fluctuations give rise to small volumes of S which will here always be taken to be spherical clusters. By definition, the free energy of bulk S is lower than that of bulk M but finite clusters also involve an interface that has higher free energy, on average, than either phase. Hence, sufficiently small clusters are typically unstable, having higher free energy even than the equivalent amount of M. Furthermore, for small clusters, there is no reason to believe that the material inside the cluster is in the bulk S state since the interface can have a volume comparable to that of the bulk region. Thus, small clusters will be unstable with respect to M: indeed, will be unstable with respect to smaller clusters thus leading to cluster dissolution. For sufficiently large clusters, the interior will consist mostly of material near the bulk S state, and the volume of the interface will be small compared to the volume of the bulk so that the cluster is stable with respect to M but unstable with respect to cluster growth thus driving the transition of the whole system from M to S [1].

In recent years, it has become apparent, first for proteins [2–6] but then even for simple fluids [4, 7, 8], that homogeneous nucleation of solids from vapor can

be more complicated than this simple picture. This is because in many cases, the conditions for solid nucleation also allow for a second metastable phase, namely that of the bulk liquid. In cases where the bulk liquid is more stable than the bulk vapor (i.e., has lower free energy) but less stable than the solid, it can be that, instead of directly nucleating a crystalline cluster from the vapor, it is energetically favorable to first nucleate a liquid cluster which then grows and, subsequently, to complete the transition to the crystalline state via a second homogeneous (liquid crystal) nucleation event within the liquid cluster (thus creating a growing solid cluster within the growing liquid cluster). This process will be referred to as "double nucleation". Whether or not this actually occurs will depend on the free energy barrier encountered in the vapor-crystal and vapor–liquid transition. Alternatively, even if direct vapor-crystal nucleation is favored, it is still possible that the growing cluster nevertheless passes through a liquid-like stage that, however, is always subcritical, as a transient on the way to creating a solid-like critical cluster. In this case, small clusters would be liquid like in structure, becoming more crystalline as they grow larger. This scenario, involving only a single nucleation event, will be referred to as "transient two-step nucleation." The simulations of ten Wolde and Frenkel [2] are clear illustrations of transient two-step nucleation: they exhibit a nucleation pathway that involves liquid-like small clusters followed by solidification as the cluster approaches the critical size. Vekilov discusses both scenarios and suggests that even when double nucleation is possible and energetically favored, it may be suppressed by kinetics [3]. Kaschiev and Vekilov have analyzed the effect of double nucleation on observed nucleation rates [9]. van Meel et al. report simulation results showing double nucleation for a Lennard–Jones fluid [7].

The key to understanding these processes is the construction of models for the free energy of inhomogeneous—multiphase—systems. Indeed, Classical nucleation theory (CNT), from which we take our lead, is fundamentally based on a description of the free energy as a function of the size of the cluster. In CNT (which should be thought of in terms of the simpler vapor–liquid transition) the width of the cluster interface is taken to be zero and the interior of the cluster is assumed to be in the bulk state, S. The only variable is then the size of the cluster and it is assumed that homogeneous nucleation consists of the growth of a cluster from size zero until it is arbitrarily large [10, 11]. This, in other words, is understood to be the "nucleation pathway." This idea can be developed more rigorously to include finite width of the interface and nonbulk properties of the interior in which case the nucleation pathway involves a description of the variation of all these properties—size, interfacial width, and interior density—simultaneously [12].

The process of solid nucleation is more complex, as the phenomenology above would imply. First, the vapor–solid transition involves at least two order parameters: the density and the *crystallinity* [4, 5, 13, 14]. Since the typical solid density is close to that of the metastable liquid (if it exists), the double nucleation scenario involves a separation of these order parameters: the vapor–liquid transition

involves two different densities but occurs at zero crystallinity while the liquid–solid transition occurs at (nearly constant) density and involves a change from zero to finite crystallinity. In the transient two-step scenario, a small (unstable) cluster begins to form at zero crystallinity but at some point, as the cluster becomes larger, the crystallinity increases so that the critical cluster has both finite density difference compared to the vapor and finite crystallinity. Both of these can be viewed as "two-step" scenarios compared to the possibility that, from the beginning, the crystallinity and density change at the same time. Clearly, description of these processes must involve a model for the free energy of inhomogeneous systems that is sufficiently detailed so as to describe liquid and solid clusters embedded in a vapor bulk. In order to capture the transient scenario, this must be supplemented by some means of determining the nucleation pathway.

Here, the basic tool for calculating the free energies will be classical, finite-temperature density functional theory (DFT). If DFT were sufficiently well developed and technically simple, nothing more would need to be said about this part of the study: unfortunately, neither of these conditions holds for the solid phase. In contrast, direct calculations for liquid–vapor systems are possible [11, 15]. DFT is only sufficiently well understood so as to give an *a priori* description of bulk fluid, bulk solid, and inhomogeneous fluid–solid systems for hard-sphere interactions. For any more realistic potential—particularly those with attractive interactions—some sort of modeling is necessary. Typically, this will involve the introduction of an effective hard-sphere diameter and the representation of the free energy as the sum of the hard-sphere functional and a second term accounting for the attractive part of the interactions. Futhermore, even for the hard-sphere functional, the calculation of the free energy for an inhomogeneous system (i.e., of a solid cluster embedded in a fluid background) is very complicated and computationally expensive [16]. Fortunately, an easier alternative is available. It has been shown that the exact free energy for a solid can be expanded in terms of gradients of the order parameters [17–19] thus providing a connection between DFT and the older gradient theories of Cahn et al. [20, 21], as well as the phase-field theories commonly used to study solid–solid transitions [22]. This is the justification for using a gradient model in the present work. A significant advantage is that the free energy functional is then only needed for the case of homogeneous (i.e., bulk) systems, thus placing less stress on the accuracy of the DFT model. On the other hand, expressions for the coefficients of the gradient terms must be found. In principle, the exact results express these in terms of derivatives of the bulk free energy but in practice, they are hard to calculate except for the case of a fluid. In this work, a semiempirical procedure will be used to fix these terms. The main difference between the present work and that of Gunton et al. [5], which is similar in spirit, is that the latter made use of toy free energy models whereas here realistic models are used that give semiquantitatively accurate bulk phase diagrams as well as of the liquid–vapor phase transition [15].

Given these approximations, the free energy for any configuration of order parameters can be calculated. The practical exploration of the models will make use of energy-surface techniques commonly applied to the study of chemical reaction pathways and structural transitions [23]. A recently developed method—involving the approximation of the spatially varying order parameters as piecewise-continuous functions, will be used to determine the critical clusters—that is, saddle points of the free energy—for homogeneous nucleation. This will already give enough information to identify if and when double nucleation is possible. The nucleation pathway will be identified, as is commonly done for chemical reaction pathways, using steepest descent paths. These identify the most likely path for the transition given the free energy surface and are a natural generalization of the CNT pathway. They do not take into account kinetic factors, such as rates of mass transport, that could play a significant role particularly for small molecules.

Section II will describe the technical details of the free energy models used. The bulk thermodynamics is used in Section III to limit the regions of the phase diagram in which double nucleation is possible. A simple model for double nucleation is also used to illustrate the role of bulk free energy differences and of surface tension. Section IV describes the application of the model to planar interfaces and illustrates the role of wetting. Detailed calculations of the energy barriers for direct nucleation of the crystal and of double nucleation are presented in Section V. The possibility of transient double nucleation is also described. The paper ends with a summary and with a discussion of future directions.

II. THE FREE ENERGY MODEL

A. Density

The central quantity defining the system state in DFT is the local density, $\rho(\mathbf{r})$. It is possible to formulate the theory with no *a priori* restrictions placed on the density, but this is computationally expensive. A commonly used alternative is to represent the density in terms of a set of basis functions. For bulk crystalline systems, this usually means a sum of Gaussians localized at the lattice sites,

$$\rho\left(\mathbf{r}; x, \rho_{\text{latt}}, \alpha\right) = x \sum_i \left(\frac{\alpha}{\pi}\right)^{3/2} \exp\left(-\alpha \left(\mathbf{r} - \mathbf{R}_i\left(\rho_{\text{latt}}\right)\right)^2\right) \tag{1}$$

where $\mathbf{R}_i$ is a lattice vector, α controls the width, x is the occupancy and ρ_{latt} is the lattice density. The average density is

$$\overline{\rho} = \frac{1}{V} \int_V \rho\left(\mathbf{r}; x, \rho_{\text{latt}}, \alpha\right) = x\rho_{\text{latt}} \tag{2}$$

The density can also be written in Fourier representation as

$$\rho\left(\mathbf{r};x,\rho_{\text{latt}},\alpha\right)=x\rho_{\text{latt}}\sum_{i}\exp\left(i\mathbf{r}\cdot\mathbf{K}_{i}\left(\rho_{\text{latt}}\right)\right)\exp\left(-\mathbf{K}_{i}^{2}\left(\rho_{\text{latt}}\right)/4\alpha\right) \tag{3}$$

where $\mathbf{K}_i$ is a reciprocal lattice vector. Notice that in this representation, it is clear that the limit $\alpha \to 0$ gives a uniform density $\rho(\mathbf{r}) = \overline{\rho}$ that describes a fluid. It is therefore natural to characterize the crystallinity by the size of the amplitude of a typical nonzero wavevector term such as

$$\chi=\exp\left(-\mathbf{K}_{1}^{2}\left(\rho_{\text{latt}}\right)/4\alpha\right)$$

so that the density can also be written as

$$\rho\left(\mathbf{r};x,\rho_{\text{latt}},\alpha\right)=x\rho_{\text{latt}}\sum_{i=0}\exp\left(i\mathbf{r}\cdot\mathbf{K}_{i}\left(\rho_{\text{latt}}\right)\right)\chi^{\left(K_{i}\left(\rho_{\text{latt}}\right)/K_{1}\left(\rho_{\text{latt}}\right)\right)^{2}} \tag{4}$$

thus showing that the density is parameterized entirely by x, ρ_{latt} and χ. In the following, we neglect the variation of the lattice density and generalize to inhomogeneous systems by allowing the occupancy and the crystallinity to depend on position. It is also convenient then to replace the occupancy by the average density so that we have

$$\rho\left(\mathbf{r}\right)=\overline{\rho}\left(\mathbf{r}\right)\sum_{i=0}\exp\left(i\mathbf{r}\cdot\mathbf{K}_{i}\left(\rho_{\text{latt}}\right)\right)\left(\chi\left(\mathbf{r}\right)\right)^{\left(K_{i}\left(\rho_{\text{latt}}\right)/K_{1}\left(\rho_{\text{latt}}\right)\right)^{2}} \tag{5}$$

The order parameters are then the local average density, $\overline{\rho}(\mathbf{r})$, and the crystallinity, $\chi(\mathbf{r})$. It will sometimes be more convenient to replace the latter by the amplitude of the smallest nonzero wavevector, $\rho_1(\mathbf{r}) = \overline{\rho}(\mathbf{r})\,\chi(\mathbf{r})$.

B. Gradient Expansion

In order to determine the density, a model for the (grand canonical) free energy functional, $\Omega[\rho]$ is necessary. Good models exist for liquids and can be used to study, for example, the liquid–vapor transition in great detail [15]. However, the theory is less well developed for the solid phase and in any case calculations for inhomogeneous solids are very expensive. The present work therefore makes use of a gradient expansion of the free energy that focuses attention on the order parameters and only requires information about homogeneous solids [17–19]. The grand-canonical free energy is written as

$$\Omega[\rho]=F[\rho]-\mu\int\rho(\mathbf{r})\,d\mathbf{r} \tag{6}$$

where μ is the chemical potential. In general, if the density can be written in terms of n order parameters, $\Gamma = \{\Gamma_1, ..., \Gamma_n\}$, as

$$\rho(\mathbf{r}) = f(\mathbf{r}; \Gamma(\mathbf{r})) \tag{7}$$

so that the density of the uniform, bulk system is

$$\rho_\Gamma(\mathbf{r}) = f(\mathbf{r}; \Gamma) \tag{8}$$

and if $\Gamma(\mathbf{r})$ in some sense "slowly varying", then the squared-gradient approximation (SGA) for the functional $F[\rho]$ is

$$F[\rho] \simeq \int \left\{ f(\Gamma(\mathbf{r})) + \frac{1}{2} K_{ab}^{ij}(\Gamma(\mathbf{r})) \frac{\partial \Gamma_i(\mathbf{r})}{\partial r_a} \frac{\partial \Gamma_j(\mathbf{r})}{\partial r_b} \right\} \tag{9}$$

where the summation convention is used. The first term of the free energy involves the bulk free energy density defined as

$$f(\Gamma) = \frac{1}{V} F[\rho_\Gamma] \tag{10}$$

For the order parameters used here, the free energy is explicitly

$$F[\rho] \simeq \int \left\{ \begin{array}{c} f(\overline{\rho}(\mathbf{r}), \chi(\mathbf{r})) + \frac{1}{2} K_{ab}^{\rho\rho}(\overline{\rho}(\mathbf{r}), \chi(\mathbf{r})) \frac{\partial \overline{\rho}(\mathbf{r})}{\partial r_a} \frac{\partial \overline{\rho}(\mathbf{r})}{\partial r_b} \\ + K_{ab}^{\rho\chi}(\overline{\rho}(\mathbf{r}), \chi(\mathbf{r})) \frac{\partial \overline{\rho}(\mathbf{r})}{\partial r_a} \frac{\partial \chi(\mathbf{r})}{\partial r_b} + \frac{1}{2} K_{ab}^{\chi\chi}(\overline{\rho}(\mathbf{r}), \chi(\mathbf{r})) \frac{\partial \chi(\mathbf{r})}{\partial r_a} \frac{\partial \chi(\mathbf{r})}{\partial r_b} \end{array} \right\} \tag{11}$$

C. Bulk Free Energy

The bulk free energy model used here is based on the idea of separating the free energy into a hard-sphere contribution, for which the DFT is well developed, and a second contribution that accounts for the long-ranged attractive interactions. A particularly simple model is based on the observation that the local structure of an FCC solid and a simple fluid are quite similar so that, as a simplest approximation, one can imagine that the correction to the hard-sphere model is independent of the local structure [4, 24] giving

$$\frac{1}{V} F[\rho_\Gamma] = \frac{1}{V} F_{\mathrm{HS}} \left[\rho_\Gamma d(\Gamma)^3 \right] + f_{\mathrm{tail}}(\rho_\Gamma, d(\Gamma)) \tag{12}$$

where $d(\Gamma) = d(\overline{\rho})$ is the effective hard-sphere diameter and

$$f_{\mathrm{tail}}(\rho_\Gamma, d(\Gamma)) = \frac{1}{V} F[\overline{\rho}] - \frac{1}{V} F_{\mathrm{HS}} \left[\overline{\rho} d^3 \right] \tag{13}$$

is the contribution of the attractive part of the interaction to the liquid free energy. The effective hard-sphere diameter is determine using the WCA expression as modified by Ree et al. [25, 26].

D. Bulk Phase Diagram

In DFT, one is always working in the grand-canonical ensemble so the external parameters are the temperature, the chemical potential, μ, and the applied external field, $\phi(\mathbf{r})$. The latter includes any confining walls: if the walls are hard, then the volume is a fixed parameter (as will always be assumed here). The appropriate free energy is the grand potential,

$$\Omega[\rho] = F[\rho] - \mu \int \rho(\mathbf{r})\, d\mathbf{r} + \int \phi(\mathbf{r})\, \rho(\mathbf{r})\, d\mathbf{r}$$

and it should be noted that all information about the state is encoded in the local density function, $\rho(\mathbf{r})$. The equilibrium states (i.e., density distributions) are determined by minimization,

$$0 = \frac{\delta \Omega[\rho]}{\delta \rho(\mathbf{r})} \tag{14}$$

which is to say

$$\frac{\delta F[\rho]}{\delta \rho(\mathbf{r})} = \mu - \phi(\mathbf{r}) \tag{15}$$

For parameterized profiles, the requirement that the free energy be a minimum gives

$$0 = \frac{\delta \Omega[\rho_\Gamma]}{\delta \Gamma(\mathbf{r})} \tag{16}$$

and if the parameters are constants, then

$$0 = \frac{\partial \Omega[\rho_\Gamma]}{\partial \Gamma} \tag{17}$$

For a given value of chemical potential, there may be multiple solutions for the density in which case the equilibrium state is the one corresponding to the absolute minimum of the grand potential. Two phase coexistence therefore requires that

$$\left.\frac{\delta F[\rho]}{\delta \rho(\mathbf{r})}\right|_{\rho_1} = \mu - \phi(\mathbf{r}) = \left.\frac{\delta F[\rho]}{\delta \rho(\mathbf{r})}\right|_{\rho_2} \tag{18}$$
$$\Omega[\rho_1] = \Omega[\rho_2]$$

In particular, using the chain rule for functional differentiation,

$$\begin{aligned}\frac{\partial F[\rho]}{\partial \overline{\rho}} &= \int \frac{\delta F[\rho]}{\delta \rho(\mathbf{r})} \frac{\partial \rho(\mathbf{r})}{\partial \overline{\rho}} d\mathbf{r} \\ &= \int \frac{\delta F[\rho]}{\delta \rho(\mathbf{r})} d\mathbf{r}\end{aligned} \tag{19}$$

gives the usual thermodynamic relation

$$\frac{\partial \frac{1}{V} F[\rho]}{\partial \overline{\rho}} = \mu - \frac{1}{V} \int \phi(\mathbf{r})\, d\mathbf{r} \tag{20}$$

Thus, for uniform densities, $\rho_i(\mathbf{r}) = \overline{\rho}_i$ and, for example, $F[\rho_i] \to F\left(\overline{\rho}_i\right)$, the conditions for coexistence are

$$\begin{aligned}\frac{\partial \frac{1}{V} F\left(\overline{\rho}_1\right)}{\partial \overline{\rho}_1} &= \mu - \frac{1}{V} \int \phi(\mathbf{r})\, d\mathbf{r} = \frac{\partial \frac{1}{V} F\left(\overline{\rho}_2\right)}{\partial \overline{\rho}_2} \\ \Omega(\rho_1) &= \Omega(\rho_2)\end{aligned} \tag{21}$$

which is to say equality of chemical potentials and of pressures since in the bulk, $\Omega = -PV$.

For the crystalline system, the parameterization used here involves not just the average density but also the crystallinity and the lattice parameter. The conditions for an extremum of the free energy are then

$$\begin{aligned}\frac{\partial \frac{1}{V} F(\overline{\rho}, \chi, \rho_{\text{latt}})}{\partial \overline{\rho}} &= \mu - \frac{1}{V} \frac{\partial}{\partial \overline{\rho}} \int \rho(\mathbf{r}; \overline{\rho}, \chi, \rho_{\text{latt}})\, \phi(\mathbf{r})\, d\mathbf{r} \\ \frac{\partial \frac{1}{V} F(\overline{\rho}, \chi, \rho_{\text{latt}})}{\partial \chi} &= -\frac{1}{V} \frac{\partial}{\partial \chi} \int \rho(\mathbf{r}; \overline{\rho}, \chi, \rho_{\text{latt}})\, \phi(\mathbf{r})\, d\mathbf{r} \\ \frac{\partial \frac{1}{V} F(\overline{\rho}, \chi, \rho_{\text{latt}})}{\partial \rho_{\text{latt}}} &= -\frac{1}{V} \frac{\partial}{\partial \rho_{\text{latt}}} \int \rho(\mathbf{r}; \overline{\rho}, \chi, \rho_{\text{latt}})\, \phi(\mathbf{r})\, d\mathbf{r}\end{aligned} \tag{22}$$

In principle, the chemical potential and external field are specified and these equations solved for $\overline{\rho}$, χ, and ρ_{latt}: in practice, it is simpler to choose a value of the lattice density and to use these conditions to determine $\overline{\rho}$, χ, and μ. Further technical details are discussed in Appendix A.

III. THERMODYNAMICS OF TWO STEP NUCLEATION

A. Independent Variables and Ensembles

The main calculational tool used here, DFT, is formulated in the grand-canonical ensemble in which the independent variables are the temperature, chemical

potential, and volume (or, more generally, applied field) and the free energy of interest is the grand potential. Experiments are typically performed at constant pressure, temperature, and volume for which the Gibbs free energy is relevant. Fortunately, in discussing nucleation, one is always interested in free energy differences and these are independent of the ensemble [27]. In the following, since it is based on DFT, the grand-canonical ensemble is always used, however in many cases the difference between ensembles can be suppressed by focusing on physical quantities. Thus, rather than specify the chemical potential (for the grand ensemble) or the pressure (for the PVT ensemble) one can specify the density of the final phase that is a meaningful variable in both formulations and in either case is uniquely related to the state variable.

B. Interaction Potentials and the Fluid Phase

In this study, the nucleation properties of two different systems will be considered: simple fluids as described by the Lennard–Jones potential,

$$v_{\mathrm{LJ}}(r) = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) \tag{23}$$

and globular proteins as described by the ten Wolde–Frenkel model interaction,

$$v_{\mathrm{tWF}}(r) = \begin{cases} \infty, & r > \sigma \\ \frac{4\epsilon}{\alpha^2}\left(\left(\frac{1}{(\frac{r}{\sigma})^2-1}\right)^6 - \alpha\left(\frac{1}{(\frac{r}{\sigma})^2-1}\right)^3\right), & r \geq \sigma \end{cases} \tag{24}$$

which will be studied here for the value $\alpha = 50$ as is appropriate to model the phase behavior of globular proteins.

Both of these potentials are long ranged in the sense of decaying as power laws. In simulation, infinite-ranged potentials are difficult to deal with, so any long-ranged potential $v(r)$ is typically cutoff at some distance, r_{c}. For Monte Carlo, a simple shift of the potential to make the energy continuous at the cutoff is typically used so that the so-called truncated and shifted potential is

$$v_{\mathrm{mc}}(r) = \begin{cases} v(r) - v(r_{\mathrm{c}}), & r \leq r_{\mathrm{c}} \\ 0, & r > r_{\mathrm{c}} \end{cases} \tag{25}$$

In molecular dynamics simulations, the force is usually required to be continuous so that the force-shifted potential is commonly used,

$$v_{\mathrm{md}}(r) = \begin{cases} v(r) - v(r_{\mathrm{c}}) - v'(r_{\mathrm{c}})(r - r_{\mathrm{c}}), & r \leq r_{\mathrm{c}} \\ 0, & r > r_{\mathrm{c}} \end{cases} \tag{26}$$

where $v'(r) = dv(r)/dr$. Other forms are also important, particularly the Broughton–Gilmer modification of the LJ potential:

$$v_{\mathrm{BG}}(r) = \begin{cases} v_{\mathrm{LJ}}(r) + C_1, \ r \le 2.3\sigma \\ C_2 \left(\frac{\sigma}{r}\right)^{12} + C_3 \left(\frac{\sigma}{r}\right)^{6} + C_4 \left(\frac{\sigma}{r}\right)^{2} + C_5, \ 2.3\sigma < r \le 2.5\sigma \\ 0, \ r > 2.5\sigma \end{cases} \tag{27}$$

where $C_1 = 0.016132\epsilon$, $C_2 = 313.66\epsilon$, $C_3 = -68.069\epsilon$, $C_4 = 0.083312\epsilon$, and $C_5 = 0.74689\epsilon$ [28].

All of these details significantly affect the liquid–vapor phase diagram. The DFT model described above requires as input both the interaction potential and the equation of state for the fluid phase. For the LJ potential with *no cutoff*, essentially exact empirical equations of state are available for temperatures above the triple point [29–31]. For finite cutoffs, these must be modified with inexact mean-field corrections. Although useful for studying liquid–vapor coexistence [11, 15], they are of limited utility for vapor–solid nucleation since the interesting affects occur below the triple point. Less accurate, but more broadly applicable, are mean-field equations of state based on thermodynamic perturbation theory such as the Weeks–Chandler–Andersen (WCA) theory [25, 26, 32–34] which will be used here.

C. Solid Phase Diagrams and Intermediate Phases

Figure 1 shows the calculated vapor–liquid–FCC solid phase diagram for the infinite-ranged LJ potential and for the truncated and shifted tWF potential compared to simulation in terms of the reduced temperature $T^* = k_{\mathrm{B}}T/\epsilon$. The LJ phase diagram possesses both a liquid–vapor critical point and a triple point whereas for the protein model, the liquid–vapor transition is metastable. Since the fluid equation of state is being calculated from thermodynamic perturbation theory, which is a mean-field theory, the critical point is, as expected, poorly described for both potentials. Apart from this expected inaccuracy, the model is in semiquantitative agreement with simulation for all three phases.

1. Nucleation Scenarios for Globular Proteins

In order to clarify the thermodynamics of metastable states, we consider in more detail the phase diagram derived from the tWF potential. Figure 2 shows a line crossing the phase diagram at constant temperature. The points A and B identify coexisting vapor and solid phases which, by definition, have the same free energy and the same control parameter (chemical potential or pressure). Starting at the vapor point A and moving along the isotherm in the direction of increasing density, that is, to the right, corresponds to increasing the chemical potential and to each value of the density there will be a point on the isotherm to the right of the

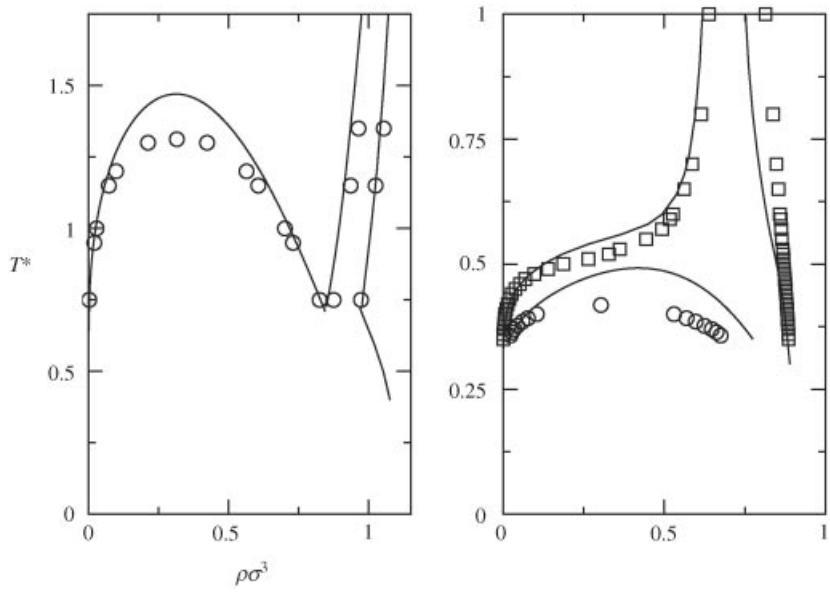


Figure 1. Phase diagrams for the Lennard–Jones potential, left panel, and the tWF potential, right panel, together with simulation data. The tWF potential was cutoff and shifted at $r = 2.8\sigma$.

solid-branch of the binodal, that is, to the right of point B, having the same chemical potential; however, under these conditions the solid phase will have lower free energy than the vapor phase so that the vapor points to the right of A are *metastable* with respect to the solid phase. Eventually, as one moves along the isotherm, the point in the vapor region reaches the vapor–liquid coexistence curve so that there is a coexisting liquid phase. However, by definition, this particular liquid state has the same free energy as the vapor, one knows that it has higher free energy than the solid and so can only be metastable. This remains true as one moves along the isotherm to densities to the right of the vapor branch of the liquid–vapor binodal. Eventually, one reaches the vapor–liquid spinodal and above this density, the vapor phase does not exist. The set of liquid points with the same chemical potential as the vapor points on the spinodal will therefore divide region to the right of the liquid binodal into a lower density region, for which a vapor with the same chemical potential can always be found, and a higher density region with no corresponding vapor.

What is the role of the metastable liquid phase in vapor–solid nucleation? We can only address this question here with respect to double nucleation. (Transient metastable states are a property of the nucleation pathway and its presence or absence cannot be answered solely from a consideration of the bulk phases.) First,

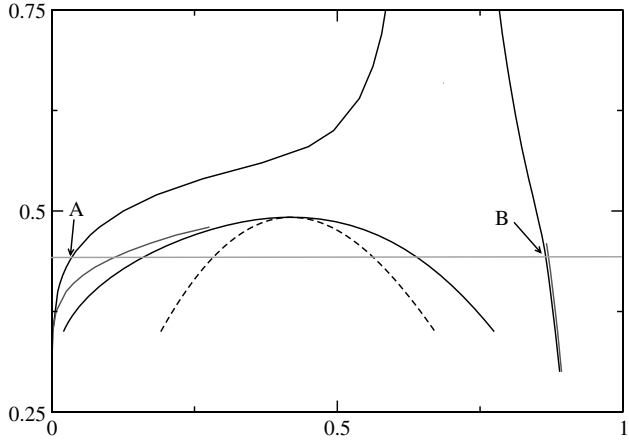


Figure 2. The phase diagram for the tWF potential as in Fig. 1 but without the simulation data and showing the spinodal for the vapor–liquid transition (broken line). The figure also shows, to the right of point A, the boundaries for the existence of the intermediate liquid phase: there is no liquid phase for chemical potentials corresponding to vapor densities to the left of the new line. A similar line for the solid phase is nearly indistinguishable from the solid binodal. Similar boundaries also exist on the Lennard–Jones phase diagram but are so close to the binodals as to be almost indistinguishable. The horizontal line, an isotherm, picks out coexisting states with the coexisting vapor and solid states being labeled A and B, respectively.

consider again the vapor–liquid coexistence curve. By a reversal of the reasoning above, vapor states to the left of the vapor branch correspond to liquid points to the left of the liquid branch of the coexistence curve and are more stable than the corresponding liquid state. Moving to the left away from the vapor branch, the corresponding liquid point also moves left until it reaches the liquid–vapor spinodal: beyond this point, there is no corresponding liquid. This therefore defines a line in the phase diagram with the property that vapor states to the left have no corresponding liquid states, shown as a red line in Fig. 2. In particular, points on the vapor branch of the vapor–solid coexistence curve have no corresponding liquid. This line therefore divides the metastable vapor region into two parts: a low-density part in which there is no corresponding liquid state (i.e., no liquid with the same chemical potential) that can play a role in nucleation and a higher density region where the corresponding liquid exists as a metastable state. Hence, *for systems prepared with the vapor density between the vapor branch of the solid–vapor coexistence curve and the new demarcation line, double nucleation is not possible as there is no metastable liquid.* The region within which vapor and liquid states with the same chemical potential can be found is therefore an envelope around the binodal and will be referred to as the μ-nodal curve.

Starting at the vapor branch of the vapor–solid coexistence curves, and moving to the right we therefore find that

1. From coexistence to the μ-nodal line, there is no liquid phase and so, double nucleation is not possible. In this region, the solid free energy is less than that of the vapor.
2. From the μ-nodal line, to the vapor branch of the vapor–liquid coexistence curves, there is a liquid state but it has higher free energy than the vapor (which has higher free energy than the solid). Nucleation via a liquid cluster is possible, with the solid cluster nucleating within the liquid but the liquid cluster would always be metastable. This is a form of transient two-step nucleation.
3. From the vapor branch of the liquid–vapor coexistence curve to the vapor–liquid spinodal, the liquid has lower free energy than the vapor but higher than the solid. True double nucleation is possible depending on the barriers for the formation of a critical liquid cluster within the vapor, a critical solid cluster in the liquid as compared to the barrier for directly forming a critical solid cluster in the vapor. Even if the latter is favored, a transient scenario is possible.

All of this is shown in a more direct way for the tWF potential in Fig. 3 where supersaturation, $S = P_v/P_{coex}$ with P_v the vapor pressure and P_{coex} the vapor–solid coexistence pressure, is used as the independent variable. By definition, for $S < 1$, the vapor is the stable phase and for $S > 1$ the solid is the stable phase. The latter region is divided into three sections: that for which there is no corresponding

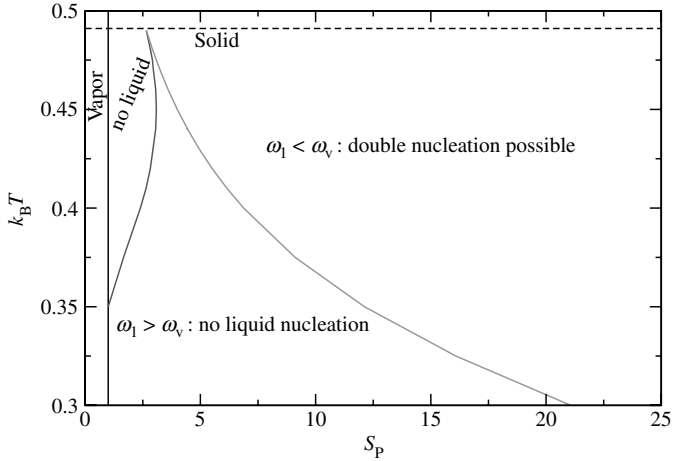


Figure 3. Phase diagrams for the tWF potential with (vapor–solid) supersaturation as the independent variable.

liquid state, that in which there is a liquid but it has higher free energy than the vapor, and that in which there is a liquid state with lower free energy than the vapor but higher free energy than the solid. Only in the latter case is liquid nucleation possible so it is only in this region of the phase diagram that double nucleation can occur. This therefore represents a set of necessary conditions for double nucleation. Nucleation pathways that do not involve double nucleation but which pass through liquid-like state, that is, transient two-step nucleation, can in principle occur for any value of $S > 1$. The primary goal of detailed DFT calculations is to determine the necessary conditions for double nucleation and to assess for what conditions, if any, transient two-step nucleation occurs.

2. *Nucleation Scenarios for Simple Fluids*

In some ways, the phase diagram for simple liquids is more complex because the various coexistence curves cross. As shown in Fig. 2, the calculations indicate that there is liquid–vapor coexistence below the triple point. However, just as in the case of the globular proteins, the vapor branch of the vapor–liquid coexistence curve is to the right of the vapor branch of the vapor–solid coexistence curves thus indicating that the coexisting liquid and vapor have higher free energy than does the solid phase (at the value of chemical potential or pressure corresponding to liquid–vapor coexistence). Hence, below the triple point, the liquid phase is again metastable just as in the case of the globular proteins. Similarly, below the triple point the liquid branch of the liquid–solid coexistence curves lies to the left of the liquid branch of the liquid–vapor coexistence curve, that is, it is in the metastable (or even unstable) region of the liquid–vapor phase diagram. So, these liquid states have higher free energy than the corresponding vapor phase and the coexisting liquid and solid phases are metastable with respect to the vapor. Taking all of this together, only the vapor–solid transition is usually drawn below the triple point, but there is a metastable liquid and a liquid–vapor coexistence curve in this region. It is this metastable liquid phase that gives rise to the possibility of two-step nucleation even for simple liquids. However, the fact that the vapor branches of the vapor–solid coexistence curves and the vapor–liquid coexistence curves are very close together makes it more difficult to display the various metastability boundaries. Figure 4 shows the phase information in terms of the supersaturation and the similarities and differences to the model protein behavior are evident. The main difference is that the intermediate liquid state exists for all values of the supersaturation leaving only the two regions distinguishing liquids which are less or more stable than the vapor phase.

D. A Simple Picture of Double Nucleation

In this section, the goal is to anticipate the following, more technical, developments to give an idea of how double nucleation can be described by something that looks

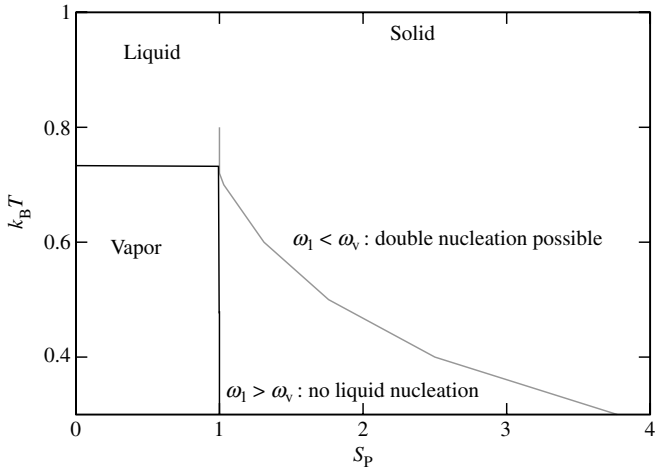


Figure 4. Phase diagram for the LJ potential with (vapor–solid) supersaturation as the independent variable.

like an extension of CNT. We begin with the free energy functional defined above specialized to spherical symmetry,

$$\Omega[\rho] \simeq \int \left\{ \omega(\overline{\rho}(r), \chi(r)) + \frac{1}{2} K^{\rho\rho} \left(\frac{\partial \overline{\rho}(r)}{\partial r} \right)^2 + K^{\rho\chi} \frac{\partial \overline{\rho}(r)}{\partial r} \frac{\partial \chi(r)}{\partial r} + \frac{1}{2} K^{\chi\chi} \left(\frac{\partial \chi(r)}{\partial r} \right)^2 \right\} \tag{28}$$

where $\omega = f - \mu\rho$ is the grand potential per unit volume and the coefficients of the gradient terms are taken to be constants. It is assumed that the temperature and chemical potential are such that there are three bulk states: a solid with order parameters $\left(\overline{\rho}_s, \chi_s\right)$, a liquid with order parameters $\left(\overline{\rho}_l, 0\right)$, and a vapor with parameters $\left(\overline{\rho}_v, 0\right)$. We now introduce a very simple model for the order parameters for a (spherically symmetric) vapor–solid cluster that nevertheless captures the important physics of the real system:

$$\overline{\rho}(\mathbf{r}) = \begin{cases} \overline{\rho}_s, & r < R \\ \overline{\rho}_s + \left(\overline{\rho}_v - \overline{\rho}_s\right) \frac{r-R}{w}, & r < R \\ \overline{\rho}_v, & r > R + w \end{cases} \tag{29}$$

$$\chi(\mathbf{r}) = \begin{cases} \chi_s, & r < R \\ \chi_s - \chi_s \frac{r-R}{w}, & r < R \\ 0, & r > R + w \end{cases}$$

This piecewise-linear model can be extended to give arbitrarily complex approximations and will play an important role below. For now, it is substituted into the expression for the free energy that is then simplified to get

$$\begin{aligned}\Omega[\rho]-\Omega_{\mathrm{v}} \simeq & \frac{4\pi}{3}R^3\left(\omega_{\mathrm{s}}-\omega_{\mathrm{v}}\right)\\ & +4\pi R^2\left(w\overline{\omega}\left(\overline{\rho}_{\mathrm{s}},\overline{\rho}_{\mathrm{v}},\chi_{\mathrm{s}}\right)+\frac{1}{2w}K^{\rho\rho}\left(\overline{\rho}_{\mathrm{s}}-\overline{\rho}_{\mathrm{v}}\right)^2\right.\\ & \left.+\frac{1}{w}K^{\rho\chi}\chi_{\mathrm{s}}\left(\overline{\rho}_{\mathrm{s}}-\overline{\rho}_{\mathrm{v}}\right)+\frac{1}{2w}K^{\chi\chi}\chi_{\mathrm{s}}^2\right)+O(Rw,w^2)\end{aligned} \tag{30}$$

with

$$\Omega_{\mathrm{v}}=\int\omega\left(\overline{\rho}_{\mathrm{v}},0\right)d\mathbf{r} \tag{31}$$

$$\overline{\omega}\left(\overline{\rho}_{\mathrm{s}},\overline{\rho}_{\mathrm{v}},\chi\right)=\int_0^1\omega\left(\overline{\rho}_{\mathrm{s}}+\lambda\left(\overline{\rho}_{\mathrm{v}}-\overline{\rho}_{\mathrm{s}}\right),(1-\lambda)\chi\right)d\lambda$$

In the present discussion, it is assumed that the clusters are not small (in the sense that $w/R\ll 1$) so that lower order terms in the expression for Ω can be neglected.

The critical cluster is a stationary point of the free energy so that the width is determined from

$$0=\overline{\omega}\left(\overline{\rho}_{\mathrm{s}},\overline{\rho}_{\mathrm{v}},\chi_{\mathrm{s}}\right)-\frac{1}{2w^2}\left(K^{\rho\rho}\left(\overline{\rho}_{\mathrm{s}}-\overline{\rho}_{\mathrm{v}}\right)^2+2K^{\rho\chi}\chi_{\mathrm{s}}\left(\overline{\rho}_{\mathrm{s}}-\overline{\rho}_{\mathrm{v}}\right)+K^{\chi\chi}\chi_{\mathrm{s}}^2\right) \tag{32}$$

and the radius from

$$\begin{aligned}0= & R\left(\omega_{\mathrm{s}}-\omega_{\mathrm{v}}\right)+2\left(w\overline{\omega}\left(\overline{\rho}_{\mathrm{s}},\overline{\rho}_{\mathrm{v}},\chi_{\mathrm{s}}\right)+\frac{1}{2w}K^{\rho\rho}\left(\overline{\rho}_{\mathrm{s}}-\overline{\rho}_{\mathrm{v}}\right)^2\right.\\ & \left.+\frac{1}{w}K^{\rho\chi}\chi_{\mathrm{s}}\left(\overline{\rho}_{\mathrm{s}}-\overline{\rho}_{\mathrm{v}}\right)+\frac{1}{2w}K^{\chi\chi}\chi_{\mathrm{s}}^2\right)\\ = & R\left(\omega_{\mathrm{s}}-\omega_{\mathrm{v}}\right)+4w\overline{\omega}\left(\overline{\rho}_{\mathrm{s}},\overline{\rho}_{\mathrm{v}},\chi_{\mathrm{s}}\right)\end{aligned} \tag{33}$$

giving

$$\Delta\Omega_{\mathrm{vs}}\simeq\frac{64\pi}{3\sqrt{2}}\frac{\overline{\omega}^{3/2}\left(\overline{\rho}_{\mathrm{s}},\overline{\rho}_{\mathrm{v}},\chi_{\mathrm{s}}\right)}{\left(\omega_{\mathrm{v}}-\omega_{\mathrm{s}}\right)^2}\left(K^{\rho\rho}\left(\overline{\rho}_{\mathrm{s}}-\overline{\rho}_{\mathrm{v}}\right)^2+2K^{\rho\chi}\chi_s\left(\overline{\rho}_{\mathrm{s}}-\overline{\rho}_{\mathrm{v}}\right)+K^{\chi\chi}\chi_{\mathrm{s}}^2\right)^{3/2} \tag{34}$$

The same model applied to the vapor–liquid and liquid–solid critical clusters gives

$$\Delta\Omega_{\rm vl} \simeq \frac{64\pi}{3\sqrt{2}} \frac{\overline{\omega}^{3/2}\left(\overline{\rho}_{\rm l}, \overline{\rho}_{\rm v}, 0\right)}{\left(\omega_{\rm v} - \omega_{\rm l}\right)^2} \left(K^{\rho\rho}\left(\overline{\rho}_{\rm l} - \overline{\rho}_{\rm v}\right)^2\right)^{3/2} \tag{35}$$

$$\Delta\Omega_{\rm ls} \simeq \frac{64\pi}{3\sqrt{2}} \frac{\overline{\omega}^{3/2}\left(\overline{\rho}_{\rm s}, \overline{\rho}_{\rm l}, \chi_{\rm s}\right)}{\left(\omega_{\rm l} - \omega_{\rm s}\right)^2} \left(K^{\rho\rho}\left(\overline{\rho}_{\rm s} - \overline{\rho}_{\rm l}\right)^2 + 2K^{\rho\chi}\chi_{\rm s}\left(\overline{\rho}_{\rm s} - \overline{\rho}_{\rm l}\right) + K^{\chi\chi}\chi_{\rm s}^2\right)^{3/2}$$

Finally, let us make the further approximation that the density of the liquid and solid states are almost the same so that

$$\Delta\Omega_{\rm vl} \simeq \frac{64\pi}{3\sqrt{2}} \frac{\overline{\omega}^{3/2}\left(\overline{\rho}_{\rm s}, \overline{\rho}_{\rm v}, 0\right)}{\left(\omega_{\rm v} - \omega_{\rm l}\right)^2} \left(K^{\rho\rho}\left(\overline{\rho}_{\rm s} - \overline{\rho}_{\rm v}\right)^2\right)^{3/2}$$

$$\Delta\Omega_{\rm ls} \simeq \frac{64\pi}{3\sqrt{2}} \frac{\overline{\omega}^{3/2}\left(\overline{\rho}_{\rm s}, \overline{\rho}_{\rm s}, \chi_{\rm s}\right)}{\left(\omega_{\rm l} - \omega_{\rm s}\right)^2} \left(K^{\chi\chi}\chi_{\rm s}^2\right)^{3/2} \tag{36}$$

Double nucleation will be possible if $\omega_{\rm s} < \omega_{\rm l} < \omega_{\rm v}$ and will be the energetically preferred pathway provided $\Delta\Omega_{\rm vs} > \Delta\Omega_{\rm vl}$ that in turn is more likely if one or more of these conditions is fulfilled:

1. The barrier for a direct transition is larger than that for an indirect transition $\overline{\omega}\left(\overline{\rho}_{\rm s}, \overline{\rho}_{\rm v}, \chi_{\rm s}\right) > \overline{\omega}\left(\overline{\rho}_{\rm l}, \overline{\rho}_{\rm v}, 0\right), \overline{\omega}\left(\overline{\rho}_{\rm l}, \overline{\rho}_{\rm v}, 0\right)$. (It was this condition that was studied previously by Lutsko and Nicolis [4]).
2. $\omega_{\rm v} - \omega_{\rm l} \gg \omega_{\rm v} - \omega_{\rm s}$
3. $K^{\rho\chi}$, $K^{\chi\chi}$ are not small compared to $K^{\rho\rho}$.

Clearly, the factor $(\omega_{\rm v} - \omega_{\rm l})^2$ occurring in the denominator of $\Delta\Omega_{\rm vl}$ can be important in raising the vapor–liquid barrier compared to the vapor–solid barrier. On the other hand, the factor involving the gradient coefficients is always going to be smaller for the vapor–liquid cluster than for the vapor–solid cluster due to the terms related to crystallinity.

IV. GRADIENT COEFFICIENTS AND PLANAR INTERFACES

In the standard CNT model, the excess free energy of a cluster is the sum of a bulk term and a surface term with the latter being proportional to the planar surface tension at coexistence. In the same way, for the model considered here, the planar surface tension will play a key role in determining the gradient coefficients.

A. Gradient Coefficients

Since the second-gradient approximation is the result of a formal expansion of the free energy, exact expressions for the gradient coefficients exist (see Appendix B).

Their evaluation requires full knowledge of the direct correlation function in the bulk system for all densities and crystallinities that is of course not known. While reasonable models could be used to make approximate evaluations of the coefficients, the results would probably be disappointing when used to evaluate physical quantities such as the liquid–solid or vapor–solid planar surface tension simply because the squared-gradient model is a crude approximation. In fact, the relevant interfaces tend to have widths of only a few times the typical atomic separation so that the assumption of slowly varying order parameters that underlies the SGA is unlikely to be very good, although for the particular case of the liquid–vapor interface that involves only a single order parameter and no structural change, it is actually rather good [12]. Thus, while it would be interesting to make good evaluations of the exact expressions for the coefficients, the approach used here is more pragmatic. First, it is noted that a systematic expansion in crystallinity and density gives

$$\begin{aligned} K^{\rho\rho}(\overline{\rho}, \chi) &= K^{\rho\rho}(0,0)(1 + O(\chi, \overline{\rho})) \\ K^{\rho\chi}(\overline{\rho}, \chi) &= C(T)\,\overline{\rho}\chi(1 + O(\chi, \overline{\rho})) \\ K^{\chi\chi}(\overline{\rho}, \chi) &= C(T)\,\overline{\rho}^2(1 + O(\chi, \overline{\rho})) \end{aligned} \tag{37}$$

with

$$C(T) \equiv \left[\frac{\partial^2 K^{\rho\chi}(\overline{\rho}, \chi)}{\partial\overline{\rho}\partial\chi}\right]_{\overline{\rho}=\chi=0} = \left[\frac{\partial^2 K^{\chi\chi}(\overline{\rho}, \chi)}{\partial\overline{\rho}^2}\right]_{\overline{\rho}=\chi=0}$$

Second, there is evidence that the density–density coefficient can be well modeled in the liquid, that is, for $\chi = 0$, by a density-independent quantity [12]. Third, explicit calculations in the accessible limit of low density indicate that all three coefficients are relatively insensitive to the crystallinity (aside from the explicit factors shown above) at least up to crystallinities about half that of the bulk solid. Finally, Laird has noted [35, 36] that the structural properties tend to be dominated by the short-range repulsion of the pair potential so that the liquid–solid surface tension can be approximated by that of the hard-sphere solid evaluated with an effective hard-sphere diameter giving the useful approximation that the excess surface free energy for a planar liquid–solid interface is

$$\gamma_{\mathrm{ls}} \approx 0.617 d^2(T)/k_{\mathrm{B}}T. \tag{38}$$

This suggests that $K^{\rho\chi}$ and $K^{\chi\chi}$ should be dominated by hard-sphere contributions, which would imply that they scale linearly with temperature. It is also consistent with the model free energy functional used here in which *all* of the dependence of the free energy on the crystallinity enters through the hard-sphere part of the free

energy. All of this suggests a simple approximation for the structural coefficients along the lines of

$$K^{\rho\chi}(\overline{\rho}, \chi) \approx C_{\overline{\rho}\chi}\overline{\rho}\chi k_{\mathrm{B}}T\sigma^3$$
$$K^{\chi\chi}(\overline{\rho}, \chi) \approx C_{\chi\chi}\overline{\rho}^2 k_{\mathrm{B}}T\sigma^6 \tag{39}$$

In a final simplification, the low-density limit gives $C_{\overline{\rho}\chi} = C_{\chi\chi}$.

B. Planar Interfaces

Figure 5 shows the result of using this model to calculate the liquid–solid, vapor–solid, and vapor–liquid surface tensions at various temperatures as compared to the available simulation data. It is evident that, for example, choosing $C_{\overline{\rho}\chi} = C_{\chi\chi}$ to give, for example, a value of the liquid–solid surface tension in agreement with a single point of either simulation or the Davidchack–Laird approximation is enough to give a reasonable description of the liquid–solid and vapor–solid surface tensions over a range of temperatures. Furthermore, reasonable values for all of the physical quantities are found for a range of choices of the coefficient so that the model is relatively robust with respect to this choice.

Figure 6 shows the planar profiles calculate using $C_{\overline{\rho}\chi} = C_{\chi\chi} = 2$ for a temperature below the triple point, one near the triple point and one above the triple

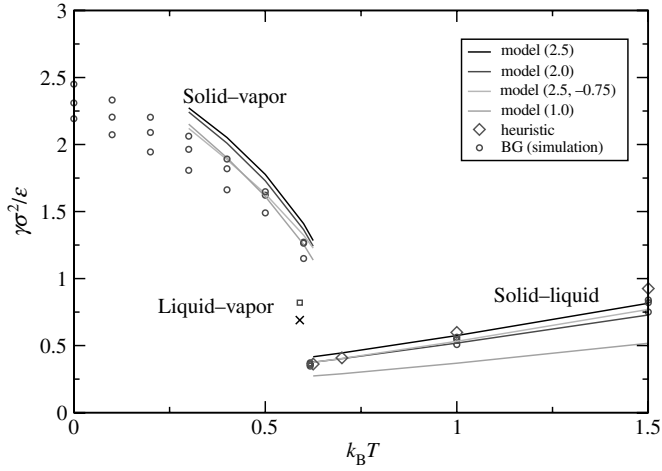


Figure 5. Excess surface free energy for the solid–vapor, liquid–vapor, and liquid–solid interfaces for a system interacting via the LJBG pair potential. The small symbols are simulation data taken from Broughton and Gilmer [37] (solid–vapor and liquid–vapor) and from Davidchack and Laird [36] (solid–liquid). The diamonds show the Laird approximation for the solid–liquid planar surface tension [35]. The simulation results at each temperature are, from highest to lowest, for the 111, 100, and 110 planes, respectively.

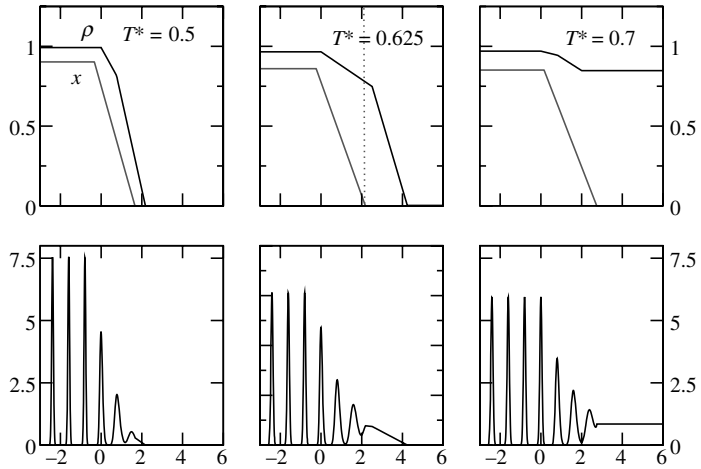


Figure 6. The order parameters (upper panels) and the 100 planar-averaged density (lower panels) for the solid–fluid interface at three temperatures that are, respectively, below, near, and above the triple point. In the upper panels, the higher curve is $\overline{\rho}(z)$ and the lower curve is $\chi(z)$. All quantities are in reduced units.

point. The profiles above and below the triple point share common features: the width and position of the transition regions for both the average density and the crystallinity are roughly the same and the overall width of the interfaces is about three atomic planes. The profile near the triple point is broader and is actually composed of two distinct regions: the first part in which there is a modest drop in density and the crystallinity goes to zero, and the second region in which the density drops to that of the vapor with the crystallinity equal to zero. The first region has the nature of a solid–liquid transition while the second has the structure of a liquid–vapor transition. In fact, this can be interpreted as a caricature of a wetted interface. True wetting, with a liquid region of finite width, is not possible in this model because the long-ranged van der Waals forces that give rise to the fluid are not being explicitly modeled. Hence, the results shown are as close as this model can come to representing wetting—basically, wetting with the bulk fluid region having zero width.

V. VAPOR-CRYSTAL NUCLEATION

A. General Considerations

Given a model for the free energy of interfacial systems it is now possible to consider the process of nucleation. In this context, nucleation means the formation of clusters that eventually become supercritical. As mentioned above, the most important issue that arises in vapor–solid nucleation is the description of the nucleation

pathway, for which there are at least three possibilities. The first is the conventional pathway in which the subcritical cluster has essentially bulk solid properties (density and crystallinity) except for an interfacial region. When the clusters are small, they are subcritical and the free energy increases with cluster size until a transition state—a critical cluster—is reached, after which further growth lowers the free energy. This pathway is therefore characterized by a single nucleation barrier and by near-bulk solid properties of the interior of the cluster. The second pathway involves double nucleation. First, a purely liquid-like cluster forms and grows until it reaches a critical size after which it is supercritical and stable with respect to the vapor. Within this liquid cluster, a second cluster forms, this time having the properties of the bulk-solid, and goes through a similar process of subcritical growth reaching a transition state and then becoming stable with respect to both the vapor and the liquid. This pathway is therefore characterized by two nucleation events and two energy barriers. The final possibility is termed transient two-step nucleation and is in some sense intermediate between the classical and double-nucleation scenarios. Sufficiently small clusters, which are of course subcritical, are liquid-like but at a certain size, below the critical size for the liquid, the crystallinity begins to increase so that there is a single critical cluster, perhaps more solid-like than liquid-like, and a single energy barrier to be crossed on the path toward crystallization. In this case, the liquid need not even be stable with respect to the vapor—indeed, there is not even *a priori* reason why the liquid must exist as a thermodynamic state (e.g., minimum of the bulk free energy) at all.

The question therefore arises as to how one uses the free energy model to characterize the nucleation pathway? Nucleation is of course a rare, noise-driven event and a dynamical description would seem most appropriate. In fact, in this sense, characterizing nucleation pathways is similar to the problem of characterizing chemical reaction pathways, for which the same issues occur.

B. Double Nucleation

The first step in characterizing any structural transition or reaction is the identification of the saddle points of the free energy surface, that is, the transition states. In general, the beginning state (the bulk vapor) and the end state (the bulk crystal) are local minima of the free energy and the transition states are the critical nuclei that define the energy barrier separating the minima. Here, it is assumed that the relevant density distributions are spherically symmetric and the order parameter fields, $\overline{\rho}(r)$ and $\chi(r)$ are again modeled by piecewise continuous functions,

$$\overline{\rho}(r) = \begin{cases} \rho_0, \ r < R \\ \rho_0 + (\rho_1 - \rho_0)\frac{r-R}{w_0}, \ R < r < R + w_0 \\ \rho_1 + (\rho_2 - \rho_1)\frac{r-R-w_0}{w_1}, \ R + w_0 < r < R + w_0 + w_1 \\ \ldots \end{cases}$$

with a similar model for the crystallinity. In order to refer to different realizations of these models, the notation $m(i, j)$ will be used indicating that there are i links for the density profile (which means $1 + 2i$ parameters since there is the initial radius and then one density and one width for each link) and j links for the crystallinity profile, giving $1 + 2j$ parameters for a total of $2 + 2i + 2j$ parameters. The free energy is then a function of those parameters and the transition states are identified by searching the $2 + 2i + 2j$ parameter space for stationary points of the free energy surface. This is done using standard eigenvector-following techniques [23].

Three specific models will be studied here. First is the "CNT" model in which there is a single link in both the density and crystallinity profiles together with the additional constraint that the radii and widths of the two profiles are the same and that the density and crystallinity in the bulk region are the same as for the bulk crystal for the given temperature and chemical potential. This model therefore depends on only two parameters (the radius and width of the profiles) and is the simplest possible model. The other models studied will be $m(1, 1)$, a single link for each profile, and $m(2, 1)$, in which there are two links in the density profile. As for the planar interfaces described above, this is necessary to allow for wetting of the surface and will be seen to lead to a substantial reduction of the free energy over the single link model. On the other hand, including additional links in the crystallinity has very little effect. (For example, at $T = 0.6$ and $S = 1.25$, the change in free energy of the critical nucleus found using the $m(2, 1)$ model and the $m(2, 2)$ model is on the order of 1%.)

The interesting feature of this problem is that for sufficiently high supersaturations, both the liquid and the solid are more stable than the vapor so that there are three transition states: one corresponding to the vapor–liquid transition, another to the vapor–solid transition, and a third to the liquid–solid transition. All of this follows simply from the bulk free energies as discussed above and illustrated in Fig. 7. The question addressed here is which of the possible paths will occur: direct transition from vapor to solid or a two-step transition via first a vapor–liquid transition and then a liquid–solid transition. It is assumed that whichever transition involves the lowest free energy barriers will dominate.

Tables I and II give the relevant free energy barriers for transitions at different supersaturations for $k_B T = 0.5$ and 0.6, respectively. It is found that for low supersaturation, corresponding to the case of large critical nuclei, the barrier for the direct vapor–solid transition is lower than that for the vapor–liquid transition. However, at larger supersaturations, the vapor–liquid transition becomes less costly thus implying that the double-nucleation scenario is favored.

Figures 8 and 9 show the structure of the critical nuclei for a few cases. In all cases, the crystallinity drops to zero before the density reaches that of the bulk vapor so, in some sense, the crystalline interior is separated from the bulk vapor by a liquid buffer. This is especially true if one considers that the structure is essentially liquid like for crystallinities less than about 0.3. However, at the

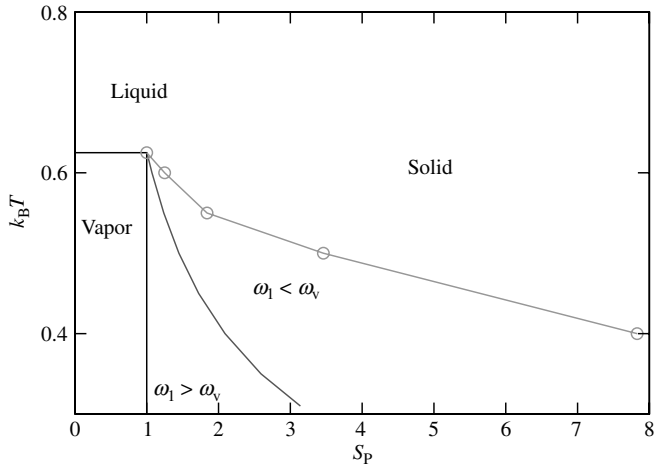


Figure 7. Phase diagram for the LJ potential with (vapor–solid) supersaturation as the independent variable. Double nucleation occurs in the region to the right and above the upper line: between the upper line and middle line, the liquid is metastable but double nucleation has a higher energy barrier than single-step nucleation of the solid.

higher temperature, a more distinctive wetting of the cluster is seen whereby the crystallinity drops to zero over a region in which the density drops from a solid-like to a liquid-like value followed by a drop of the density from liquid to vapor values. This is analogous to the planar wetting illustrated above and shows that this is also a feature of the critical clusters near the triple point.

TABLE I

Properties of Critical Nuclei for the LJBG Interaction as a Function of Supersaturation at $k_BT = 0.5$

S	2.08			3.46			5.92		
ρ_{latt}	1.00			1.005			1.01		
Model	$\beta\Omega_{VL}$	$\beta\Omega_{SV}$	$\beta\Omega_{LS}$	$\beta\Omega_{VL}$	$\beta\Omega_{SV}$	$\beta\Omega_{LS}$	$\beta\Omega_{VL}$	$\beta\Omega_{VS}$	$\beta\Omega_{LS}$
CNT	2063	1738	78	356	587	66	135	269	57
$m(1,1)$	2053	1440	73	355	463	62	133	*	53
$m(2,1)$	1899	1304	72	335	*	61	128	*	53
$\Delta\beta\Omega_{Bulk}$	−0.32	−0.74	−0.41	−0.78	−1.24	−0.46	−1.27	−1.78	−0.51

The supersaturation and lattice density are given in the first two rows followed by the excess free energy for the nuclei for vapor–liquid (VL), vapor–solid (VS), and liquid–solid (LS) nucleation. The last row gives the difference in bulk free energies for the various transitions. Blank entries occur for cases where it was not possible to find a transition state.

TABLE II
Properties of Critical Nuclei for the LJBG Interaction as a Function of Supersaturation at $k_B T = 0.6$

S	1.26			1.76			3.68		
ρ_{latt}	0.975			0.98			0.99		
Model	Ω_{LV}	Ω_{SV}	Ω_{LS}	Ω_{LV}	Ω_{SV}	Ω_{LS}	Ω_{LV}	Ω_{SV}	Ω_{LS}
CNT	4045	7013	758	396	1142	407	59	189	184
$m(1,1)$	4033	5171	708	392	783	379	58	*	171
$m(2,1)$	3714	4176	704	368	*	377	57	*	170
$\beta\Delta\Omega_{\text{Bulk}}$	−0.128	−0.227	−0.099	−0.408	−0.547	−0.139	−1.029	−1.252	−0.223

C. Transient Liquid State

Even in regions where double nucleation is not possible, the metastable liquid state could still play a role in nucleation. This becomes a question of the nucleation pathway that requires much more information than just the properties of the critical nucleus. A full description of the pathway would necessarily have to be dynamical in nature accounting for a variety of kinetic effects including heat and

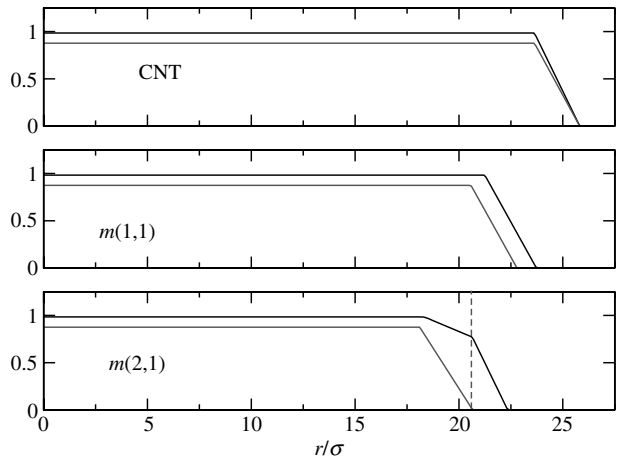


Figure 8. The structure of the critical nucleus for $T^* = 0.6$ and $S = 1.25$ as determined using the CNT, $m(1,1)$ and $m(2,1)$ parameterizations of the profiles. In each figure, the upper curve is the average density and the lower curve is the crystallinity. In the figure for the $m(2,1)$ model, a dashed line is drawn at the radius where the crystallinity becomes zero: the fact that the second link of the density profile begins at nearly this point is a clear illustration of the role of the intermediate liquid state in wetting the cluster.

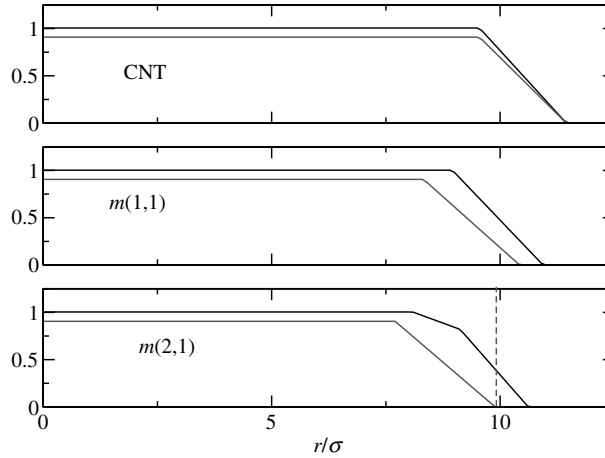


Figure 9. The same as Fig. 8 for $T^* = 0.5$ and $S = 2.08$. Desipite the existence of a metastable intermediate liquid, there is no significant wetting behavior in this case.

mass transport. It goes without saying that such a detailed description, while highly desirable, can be expected to be very difficult to implement.

Alternatively, one might imagine beginning at the transition state that is a saddle point of the free energy functional and is characterized by a size (in numbers of atoms), N_c. It would be natural then to find other stationary points under the constraint of fixed number of atoms in the cluster and thus to trace a path from the critical cluster to the bulk vapor. This is in fact similar to the methods used in simulations [2]. However, even in the case of the liquid–vapor transition this gives discontinuous paths whereby the structure changes discontinuously for some value of N. For the liquid–vapor transition, this is a transition from a well-defined cluster with a liquid-like central density to a much larger structure with a central density slightly larger than the vapor. In the present case, the transition observed is from a well-defined crystalline cluster to one with zero crystallinity.

An alternative, widely used in quantum chemistry and in the study of clusters, is the construction of steepest descent pathways away from the transition state. These are expected to be approximations to the dynamical pathway, especially in the case of quasiequilibrium dynamics due to some sort of damping: an example might be the behavior of colloids (which can often be modeled as simple fluids) or macromolecules in solution. Many different methods are used to construct the steepest descent paths (also commonly called the minimum free energy pathway or MFEP) including heuristic methods such as the nudged elastic band and the string method. Both methods have been applied to nucleation problems [10, 15] but here I expand on recent work that seeks to construct the exact MFEP for parameterized density profiles [12].

All methods of constructing steepest descent paths require the notion of a distance between two points in parameter space since "steepest descent" literally

means the path for which the energy varies most rapidly per unit distance in parameter space. The problem is that the various parameters used here—densities, widths, crystallinity—are incommensurate. However, since they exist only to specify a density profile, a natural solution is to define a metric in density–space and then to use this to induce a metric in parameter space. In the study of fluids, the metric was taken to be the Euclidean distance between two density profiles,

$$d\left[\rho, \rho'\right]^2 = \int \left(\rho(\mathbf{r}) - \rho'(\mathbf{r})\right)^2 d\mathbf{r} \tag{40}$$

It would be natural to continue to use this definition however, it is unsuitable for two reasons. The first is simply that it is technically difficult to evaluate. The second, more importantly, is that it leads to the metric being sensitive to details of the lattice structure. For example, as the radius parameter varies there is little variation in the metric until the radius crosses an atomic shell at which point there is very rapid variation. This is unacceptable in the present context since the free energy surface constructed above is based on a separation of length scales according to which the order parameters vary slowly over atomic length scales.

In order to motivate a simple alternative more in keeping with the present approach, note that for the liquid–vapor transition the crystallinity is identically zero, $\chi(\mathbf{r}) = 0$, so that the Euclidean distance function becomes

$$d\left[\rho, \rho'\right]^2 = \int \left(\overline{\rho}(\mathbf{r}) - \overline{\rho}'(\mathbf{r})\right)^2 d\mathbf{r} \tag{41}$$

which is the Euclidean distance between the $K = 0$ amplitudes in the expansion of the density. In fact, the parameterization of the density used here is

$$\rho(\mathbf{r}) = \overline{\rho}(\mathbf{r}) + \sum_{j \epsilon NN} \overline{\rho}(\mathbf{r})\, \chi(\mathbf{r})\, e^{i\mathbf{K}_j \cdot \mathbf{r}} + \cdots \tag{42}$$

where the first sum is over the first (nearest-neighbor) shell in wavevector space. Hence, the quantity $\overline{\rho}(\mathbf{r})\, \chi(\mathbf{r})$ is the spatially varying amplitude of the smallest nonzero wavevector. It therefore seems reasonable to treat this on a par with the amplitude of the zero-wavevector component and to define a distance function as

$$d\left[\rho, \rho'\right]^2 = \int \left(\overline{\rho}(\mathbf{r}) - \overline{\rho}'(\mathbf{r})\right)^2 d\mathbf{r} + \int \left(\overline{\rho}(\mathbf{r})\, \chi(\mathbf{r}) - \overline{\rho}'(\mathbf{r})\, \chi'(\mathbf{r})\right)^2 d\mathbf{r} \tag{43}$$

which is what will be used henceforth. When the order parameters are expressed in terms of a collection of scalar parameters as $\overline{\rho}(\mathbf{r}) = \overline{\rho}(\mathbf{r};\Gamma)$ and $\chi(\mathbf{r}) = \chi(\mathbf{r};\Gamma)$

this then defines a distance between points in parameter space

$$d\left[\Gamma,\Gamma'\right]^2 = \int \left(\overline{\rho}\left(\mathbf{r};\Gamma\right) - \overline{\rho}\left(\mathbf{r};\Gamma'\right)\right)^2 d\mathbf{r} + \int \left(\overline{\rho}\left(\mathbf{r};\Gamma\right)\chi\left(\mathbf{r};\Gamma\right) - \overline{\rho}\left(\mathbf{r};\Gamma'\right)\chi\left(\mathbf{r};\Gamma'\right)\right)^2 d\mathbf{r} \tag{44}$$

Assuming this function is sufficiently continuous, the distance function implies a metric

$$g_{ij}\left(\Gamma\right) = \int \left(\frac{\partial\overline{\rho}\left(\mathbf{r};\Gamma\right)}{\partial\Gamma_i}\frac{\partial\overline{\rho}\left(\mathbf{r};\Gamma\right)}{\partial\Gamma_j} + \frac{\partial\overline{\rho}\left(\mathbf{r};\Gamma\right)\chi\left(\mathbf{r};\Gamma\right)}{\partial\Gamma_i}\frac{\partial\overline{\rho}\left(\mathbf{r};\Gamma\right)\chi\left(\mathbf{r};\Gamma\right)}{\partial\Gamma_j}\right) d\mathbf{r} \tag{45}$$

The steepest descent paths are then determined by

$$\frac{d\Gamma_i}{ds} = -\frac{g_{ij}\left(\Gamma\right)\frac{\partial\beta\Omega}{\partial\Gamma_j}}{\frac{\partial\beta\Omega}{\partial\Gamma_a}g_{ab}\left(\Gamma\right)\frac{\partial\beta\Omega}{\partial\Gamma_b}} \tag{46}$$

This equation is obviously similar to a dynamics consisting of simple relaxation driven by free energy gradients as discussed in Appendix C. However, since nucleation is, fundamentally, a noise-driven process, there is no reason to expect *a priori* that the correct path can be determined by running a deterministic dynamics backward from the transition state. The idea of steepest descent is that it includes the idea of being a most probable path since it is in some sense the most efficient path over the barrier. (An analogy would be a multidimensional random walker in a potential field. In order for the walker to pass over a barrier, it must take some number of improbable steps in the right direction until it reaches the top of the barrier. The steepest descent path is the one involving the fewest number of steps. Other paths must cross the same barrier, but by including more steps, say in the "wrong" direction, the probability of falling backward toward the local minimum increases.)

The nucleation pathways have been calculated using this model for values of the supersaturation such that the metastable liquid is more stable than the vapor, but below the double-nucleation threshold. Figure 10 shows the pathway for $k_BT = 0.5$ and $S = 2.08$ plotted as a function of the total number of atoms in the cluster. It is clear that both the central density and central crystallinity begin at the values of the bulk vapor. This is because the surface tension depends on the gradient of these quantities and for very small clusters, the dominance of the surface tension term over the bulk free energy contribution forces the gradients to be zero. As the cluster grows, the density increases very rapidly while the crystallinity increases more slowly. The core of the cluster therefore densifies more quickly than it crystallizes but the effect is minor. Figure 11 shows the same quantities for $k_{\mathrm{B}}T = 0.6$, close to the triple point, and for $S = 1.25$. In this case, transient two-step nucleation is

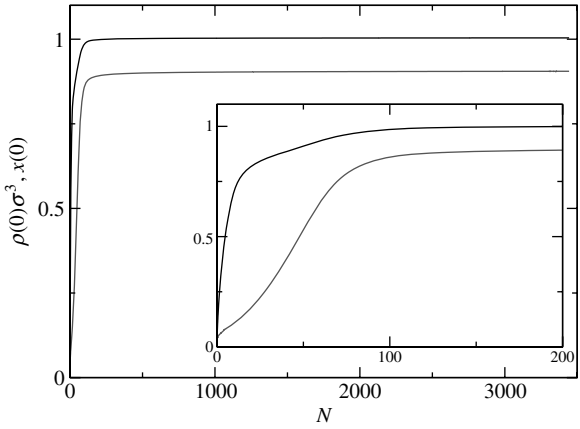


Figure 10. The average density (upper curve) and crystallinity (lower curve) at the center of the cluster as a function of cluster size for $T^* = 0.5$. The inset shows an expanded view of the early stages of nucleation.

clearly in evidence: the density increases while the crystallinity remains almost at zero until the cluster consists of well over 100 atoms. These contrasting behaviors are compared in Fig. 12 where the number of "crystalline" atoms, defined as

$$N_{\text{crys}} = \frac{1}{\chi_{\text{bulk}}} \int \overline{\rho}\,(\mathbf{r})\,\chi\,(\mathbf{r})\,d\mathbf{r} \tag{47}$$

is plotted as a function of N. Near the triple point, the number of crystalline atoms does not increase appreciably until the cluster is over 100 atoms in size

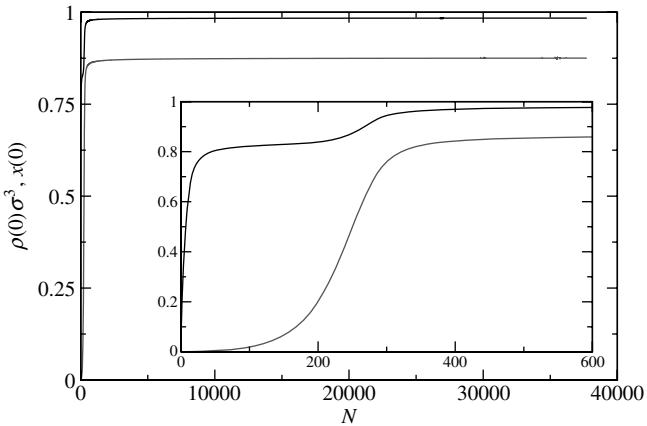


Figure 11. The average density (upper curve) and crystallinity (lower curve) at the center of the cluster as a function of cluster size for $T^* = 0.6$. The inset shows an expanded view of the early stages of nucleation.

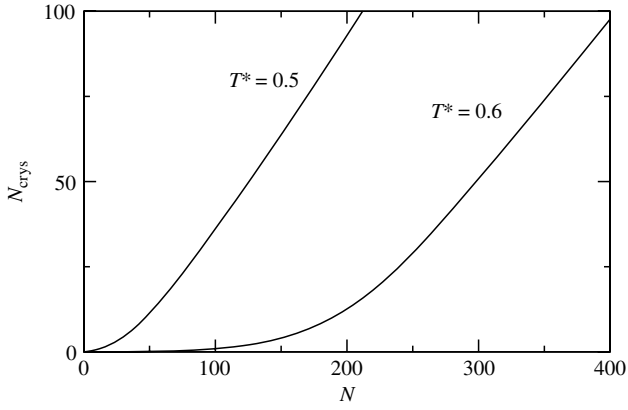


Figure 12. The number of crystalline atoms as a function of the total number of atoms in the cluster showing transient two-step nucleation near the triple point, $T^* = 0.6$, and almost completely one-step nucleation away from the triple point.

whereas it increases almost immediately at the lower temperature. This behavior is, incidentally, quite similar to the observations of ten Wolde and Frenkel who noted that for their model of globular proteins, the favored nucleation path involved two-step nucleation near the triple point but that this was not seen further below the triple point [2].

VI. CONCLUSIONS

The primary goal of this work has been they study of different pathways for the homogeneous nucleation of crystals from solution based on mean-field, DFT models. Simply from the bulk thermodynamics, it is possible to divide the phase diagram into regions for which double nucleation is and is not possible. When the bulk models are extended to inhomogeneous, multiphase systems—here via the squared-gradient approximation—it becomes possible to determine when double nucleation is more energetically favorable than single-step nucleation. Finally, by studying steepest-descent pathways connecting the transition states to the bulk states it was possible to illustrate transient two-step nucleation for the Lennard–Jones fluid. A summary of these investigations would be that

1. there is a lower supersaturation limit for double nucleation: at the triple point, the limit goes to one and it increases rapidly as the temperature is lowered;
2. transient two-step nucleation seems to be closely tied to wetting and to therefore occur most distinctly near the triple point.

Almost all phases of this study could be improved. The underlying DFT model is highly simplified and in particular the inclusion of a long-ranged van der Waals

term would give a more realistic description of wetting including a finite wetting-layer thickness. The modeling of the gradient coefficients is crude and quite empirical. One can estimate them directly from the DFT model [13, 17–19] and this might give a more realistic dependence on density and crystallinity. The piecewise continuous models could be used with more than the minimal number of links used here or some other, perhaps more physical, basis functions could be chosen (or the Euler–Lagrange equations could be solved directly without parameterizing the fields). Finally, instead of the minimum free energy pathways studied here, a more physically meaningful approach would be to search for the most *probable* pathways that requires the introduction of dynamics and noise, but which seems quite feasible based on the approach of Heymann and Vanden-Eijnden [38].

Acknowledgments

I am grateful to Gregoire Nicolis for reading an early draft of this manuscript and for a number of useful suggestions. This work was supported in part by the European Space Agency under contract number ESA AO-2004-070.

APPENDIX A: BULK SOLID PROPERTIES

The calculation of bulk thermodynamic properties for the hard-sphere solid using the given parameterization is complex. The reason is that the hard-sphere free energy functional diverges for $\overline{\rho}$ slightly greater than ρ_{latt}—at typical solid densities, the divergence occurs for $\overline{\rho} - \rho_{\text{latt}} \sim 10^{-10}$. It might seem that this is no problem as one simply takes $\overline{\rho} = \rho_{\text{latt}}$, however, upon closer inspection, this does not work.

To explain the problem, let us consider the "correct" way to fix these two densities. In principle, an equilibrium state must satisfy three conditions:

$$\left.\frac{\partial F\left[\rho_\Gamma\right]}{\partial \overline{\rho}}\right|_{\rho_{\text{latt}},\alpha} = \mu \tag{A1}$$

$$\left.\frac{\partial F\left[\rho_\Gamma\right]}{\partial \rho_{\text{latt}}}\right|_{\overline{\rho},\alpha} = 0$$

$$\left.\frac{\partial F\left[\rho_\Gamma\right]}{\partial \alpha}\right|_{\rho_{\text{latt}},\overline{\rho}} = 0$$

Let the solution to these equations be $\overline{\rho}^*$, ρ^*_{latt}, α^*. Solution of these three simultaneous equations is delicate due to the fact that (a) $|\overline{\rho} - \rho_{\text{latt}}| \ll 1$ and (b) the surprising fact that

$$\left|\left.\frac{\partial F\left[\rho_\Gamma\right]}{\partial \overline{\rho}}\right|_{\overline{\rho}=\rho^*_{\text{latt}},\rho^*_{\text{latt}},\alpha^*} - \left.\frac{\partial F\left[\rho_\Gamma\right]}{\partial \overline{\rho}}\right|_{\overline{\rho}^*,\rho^*_{\text{latt}},\alpha^*}\right| = O(\mu) \tag{A2}$$

In words, the free energy is a *very* rapidly varying function of $\overline{\rho}$ for $\overline{\rho}$ near ρ_{latt}.

This situation is somewhat better than it seems since in practical calculations, we are almost always performing some sort of search over, or tabulation in terms of, the chemical potential. Hence, rather than being given μ and having to solve three simultaneous equations, we can often take, for example, ρ_{latt} as the independent parameter and search for $\overline{\rho}^*$, α^* satisfying

$$\left.\frac{\partial F\left[\rho_\Gamma\right]}{\partial \rho_{\text{latt}}}\right|_{\overline{\rho}^*,\rho_{\text{latt}},\alpha^*} = 0 \qquad \text{(A3)}$$

$$\left.\frac{\partial F\left[\rho_\Gamma\right]}{\partial \alpha}\right|_{\overline{\rho}^*,\rho_{\text{latt}},\alpha^*} = 0$$

We then calculate μ given these values. (Note however that even evaluating the various derivatives numerically is quite delicate when the range over which $\overline{\rho}d^3$ can be varied symmetrically might be 10^{-10} or smaller.)

There is a heuristic that alleviates many of these problems. Intuitively, one tends to think in terms of *the* density and not to distinguish between the lattice density and the average density. In fact, in most calculations for the hard-sphere solid, the approximation $\overline{\rho} = \rho_{\text{latt}}$ is made and the results appear quite reasonable. Since

$$\frac{df\left(z,z\right)}{dz} = \left.\frac{\partial f\left(x,y\right)}{\partial x}\right|_{x=y=z} + \left.\frac{\partial f\left(x,y\right)}{\partial y}\right|_{x=y=z} \qquad \text{(A4)}$$

this suggests that it must be the case that

$$\left.\frac{\partial F\left[\rho_\Gamma\right]}{\partial \overline{\rho}}\right|_{\overline{\rho}=\rho^*_{\text{latt}},\rho^*_{\text{latt}},\alpha^*} + \left.\frac{\partial F\left[\rho_\Gamma\right]}{\partial \rho_{\text{latt}}}\right|_{\overline{\rho}=\rho^*_{\text{latt}},\rho^*_{\text{latt}},\alpha^*} \simeq \left.\frac{\partial F\left[\rho_\Gamma\right]}{\partial \overline{\rho}}\right|_{\overline{\rho}^*,\rho^*_{\text{latt}},\alpha^*} = \mu \qquad \text{(A5)}$$

In fact, this suspicion is borne out in practice. Combining these two points, and noting that the value of α that makes the free energy stationary varies slowly as a function of the various densities, an efficient practical procedure is to choose a value of ρ_{latt}, to set $\overline{\rho} = \rho_{\text{latt}}$, to determine the stationary value of α and to use the left-hand side of Eq. (A5) to evaluate μ. This involves a simple, controlled minimization and no need to evaluate the free energy near the divergence. A proof that Eq. (A5) is exact, or at least a good approximation, is to my knowledge missing and would be useful.

Finally, an important point is that these technicalities play no role in the calculations presented here. Since we deal here with interfacial problems, the important thing is how the free energy varies as the average density varies from that of the solid down to a relatively low value (that of a liquid or vapor). Furthermore, in clusters, the interior density is always somewhat below that of the bulk solid (dwarfing the comparatively tiny difference between $\overline{\rho}^*$ and ρ^*_{latt}) . The only way these technical points would be important if we tried to do a free minimization of the free energy for a planar interface in which case the algorithm would have to

find the correct values for the bulk system. This would essentially mean solving Eq. (A3) that, because of the divergences very near the correct solution, leads to numerical challenges. However, using the piecewise-continuous approximation, we always set the bulk values "by hand" and thereby avoid this problem.

APPENDIX B: FORMAL RESULTS FOR GRADIENT COEFFICIENTS

They are given by

$$\beta K_{ab}^{\Gamma_a\Gamma_b}(\Gamma) = -\frac{1}{2V}\int \mathbf{r}_{12,a}\mathbf{r}_{12,b}c(\mathbf{r}_1,\mathbf{r}_2;\Gamma)\frac{\partial\rho(\mathbf{r}_1;\Gamma)}{\partial\Gamma_a}\frac{\partial\rho(\mathbf{r}_2;\Gamma)}{\partial\Gamma_b}d\mathbf{r}_1 d\mathbf{r}_2$$

where $c(\mathbf{r}_1,\mathbf{r}_2;\Gamma)$ is the direct correlation function for the bulk solid. Unfortunately, the DFT models used here do not give realistic expressions for this quantity although some useful results are possible. In particular, expanding in the crystallinity gives

$$\begin{aligned} K^{\rho\rho} &= I_1(\overline{\rho}) + O(\chi) \\ K^{\rho\chi} &= \overline{\rho}\chi I_2(\overline{\rho}) + O\left(\chi^2\right) \\ K^{\chi\chi} &= \overline{\rho}^2 I_2(\overline{\rho}) + O(\chi) \end{aligned} \tag{B1}$$

with

$$\begin{aligned} I_1(\overline{\rho}) &= \frac{2\pi}{3}\int_0^\infty c(r;\overline{\rho})\,r^4 dr \\ I_2(\overline{\rho}) &= \frac{2\pi}{3}N_1\int_0^\infty \frac{\sin(K_1 r)}{K_1}c(r;\overline{\rho})\,r^3 dr \end{aligned} \tag{B2}$$

which only requires the DCF in the bulk fluid (albeit, at solid densities). Similarly, expanding in density gives

$$\beta K_{ab}^{\Gamma_a\Gamma_b}(\Gamma) = -\frac{1}{2V}\int \mathbf{r}_{12,a}\mathbf{r}_{12,b}\left(1-e^{-\beta v(r)}\right)\frac{\partial\rho(\mathbf{r}_1;\Gamma)}{\partial\Gamma_a}\frac{\partial\rho(\mathbf{r}_2;\Gamma)}{\partial\Gamma_b}d\mathbf{r}_1 d\mathbf{r}_2 + \cdots \tag{B3}$$

and combining the two expansions gives

$$\begin{aligned} K^{\rho\rho} &= I_1(0)(1 + O(\chi,\overline{\rho})) \\ K^{\rho\chi} &= \overline{\rho}\chi I_2(0)(1 + O(\chi,\overline{\rho})) \\ K^{\chi\chi} &= \overline{\rho}^2 I_2(0)(1 + O(\chi,\overline{\rho})) \end{aligned} \tag{B4}$$

and even if they did, the coefficients evaluated from these expressions might well give poor values of the surface tension due to the truncation of the gradient expansion.

Coexistence occurs at a given temperature for a single value of the chemical potential corresponding, by definition, to supersaturation equal to 1.

APPENDIX C: STEEPEST DESCENT AND DYNAMICS

Typical dynamical models depend on a distinction between order parameters that are densities of conserved quantities (such as the total mass in the canonical ensemble) and those that are densities of nonconserved quantities (such as the mass in the grand canonical ensemble). We begin with the latter case of a nonconserved order parameter. Then, the evolution is often assumed to be of the form

$$\frac{d\psi_t(\mathbf{r})}{dt} = -\Gamma \frac{\delta \Omega[\psi_t]}{\delta \psi_t(\mathbf{r})}$$

where Γ is a transport coefficient. If space is discretized and we denote $\psi_t(\mathbf{r}_i) = \psi_{ti}$, this takes the form

$$\frac{d\psi_{ti}}{dt} = -\Gamma \frac{\partial \Omega[\psi_t]}{\partial \psi_{ti}}$$

Here, the notation indicates that Ω is a function of all of the order parameters $\{\psi_{ti}\}$. Let us suppose that the system is described by some alternate set of order parameters, $\{\phi_{ti}\}$, which is complete in the sense that the two sets are equivalent and the relation between them is invertible: $\phi_{ti} = \phi_{ti}[\psi]$ and $\psi_{ti} = \psi_{ti}[\phi]$. Then, it follows that

$$\begin{aligned}\frac{d\phi_{ti}}{dt} &= -\Gamma \sum_j \frac{\partial \Omega[\psi_t]}{\partial \psi_{tj}} \frac{\partial \phi_{ti}}{\partial \psi_{tj}} \\ &= -\Gamma \frac{\partial \Omega[\phi_t]}{\partial \phi_{t1}} \sum_j \frac{\partial \phi_{t1}}{\partial \psi_{tj}} \frac{\partial \phi_{ti}}{\partial \psi_{tj}}\end{aligned} \tag{C1}$$

Suppose we did not know about this dynamics and simply wanted to write down the steepest descent equations for these models. In that case, it is necessary to specify a metric. We assume that the ψ-space is Euclidean so that the distance between two sets of fields is

$$\begin{aligned}d^2\left[\psi_t(\mathbf{r}), \psi'_t(\mathbf{r})\right] &= \int \left(\psi_t(\mathbf{r}) - \psi'_t(\mathbf{r})\right)^2 d\mathbf{r} \\ &\to \sum_i \left(\psi_{ti} - \psi'_{ti}\right)^2\end{aligned} \tag{C2}$$

This implies that

$$d^2\left[\phi_t(\mathbf{r}),\phi_t'(\mathbf{r})\right]\rightarrow\sum_i\left(\psi_{ti}\left[\phi_t\right]-\psi_{ti}\left[\phi_t'\right]\right)^2 \tag{C3}$$

Assuming sufficient analyticity, this prescribes a Riemann geometry with metric

$$g_{\mathrm{lm}}=\sum_i\frac{\partial\psi_{ti}\left[\phi_t\right]}{\partial\phi_{t\mathrm{l}}}\frac{\partial\psi_{ti}\left[\phi_t\right]}{\partial\phi_{t\mathrm{m}}} \tag{C4}$$

In general, the steepest descent equations are

$$\frac{d\phi_{ti}}{ds}=-\frac{g^{i\mathrm{l}}\dfrac{\partial\Omega\left[\phi_t\right]}{\partial\phi_{t\mathrm{l}}}}{\left(\dfrac{\partial\Omega\left[\phi_t\right]}{\partial\phi_{ta}}g^{ab}\dfrac{\partial\Omega\left[\phi_t\right]}{\partial\phi_{tb}}\right)} \tag{C5}$$

Comparison of this to Eq. (C1) shows that the dynamics is equivalent to steepest descent with the relation between time and the distance parameter being

$$\Gamma dt=\frac{ds}{\left(\dfrac{\partial\Omega\left[\phi_t\right]}{\partial\phi_{ta}}g^{ab}\dfrac{\partial\Omega\left[\phi_t\right]}{\partial\phi_{tb}}\right)} \tag{C6}$$

References

1. D. Kashchiev, *Nucleation : Basic Theory with Applications,* Butterworth-Heinemann, Oxford, 2000.
2. P. R. ten Wolde and D. Frenkel, *Science*, **77**, 1975 (1997).
3. P. G. Vekilov, *Cryst. Growth Design*, **4**, 671 (2004).
4. J. F. Lutsko and G. Nicolis, *Phys. Rev. Lett.*, **96**, 046102 (2006)
5. A. Shiryayev and J. D. Gunton, *J. Chem. Phys.*, **120**, 8318 (2004)
6. J. D. Gunton, A. Shiryayev, and D. L. Pagan, *Protein Condensation: Kinetic Pathways to Crystallization and Disease,* Cambridge University Press, Cambridge, 2007.
7. J. A. van Meel, A. J. Page, R. P. Sear, and D. Frenkel, *J. Chem. Phys.*, **129**, 204505 (2008).
8. B. Chen, H. Kim, S. J. Keasler, and R. B. Nellas, *J. Phys. Chem. B*, **112**, 4067 (2008).
9. D. Kashchiev, P. G. Vekilov, and A. B. Kolomeisky, *J. Chem. Phys*, **122**, 244706 (2005).
10. J. F. Lutsko, *Europhys. Lett.*, **83**, 46007 (2008).
11. J. F. Lutsko, *J. Chem. Phys.*, **129**, 244501 (2008).
12. J. F. Lutsko, Density functional theory of inhomogeneous liquids IV. Squared-gradient approximation and classical nucleation theory (2010), unpublished.
13. R. Ohnesorge, H. Löwen, and H. Wagner, *Phys. Rev. A*, **43**, 2870 (1991).
14. V. Talanquer and D. W. Oxtoby, *J. Chem. Phys.*, **109**, 223 (1998).
15. J. F. Lutsko, *J. Chem. Phys.*, **128**, 184711 (2008).

16. V. B. Warshavsky and X. Song, *Phys. Rev. E*, **73**, 031110 (2006).
17. H. Löwen, T. Beier, and H. Wagner, *Europhys. Lett.*, **9**, 791 (1989).
18. H. Löwen, T. Beier, and H. Wagner, *Z. Phys. B*, **79**, 109 (1990).
19. J. F. Lutsko, *Physica A*, **366**, 229 (2006).
20. J. W. Cahn and J. E. Hilliard, *J. Chem. Phys.*, **28**, 258 (1958).
21. J. W. Cahn and J. E. Hilliard, *J. Chem. Phys.*, **31**, 688 (1959).
22. H. Emmerich, *Adv. Phys.*, **57**, 1 (2008).
23. D. Wales, *Energy Landscapes,* Cambridge University Press, Cambridge, 2003.
24. W. A. Curtin, *Phys. Rev. B*, **39**, 6775 (1989).
25. F. H. Ree, *J. Chem. Phys.*, **64**, 4601 (1976).
26. H. S. Kang, C. S. Lee, T. Ree, and F. H. Ree, *J. Chem. Phys.*, **82**, 414 (1985).
27. D. W. Oxtoby and R. Evans, *J. Chem. Phys.*, **89**, 7521 (1988).
28. J. Q. Broughton and G. H. Gilmer, *J. Chem. Phys.*, **79**, 5095 (1983).
29. J. K. Johnson, J. A. Zollweg, and K. E. Gubbins, *Mol. Phys.*, **78**, 591 (1993).
30. M. Mecke, A. Müller, J. Winkelmann, J. Vrabec, J. Fischer, R. Span, and W. Wagner, *Int. J. Thermophys.*, **17**, 391 (1996).
31. M. Mecke, A. Müller, J. Winkelmann, J. Vrabec, J. Fischer, R. Span, and W. Wagner, *Int. J. Thermophys.* **19**, 1493 (1998).
32. D. Chandler and J. D. Weeks, *Phys. Rev. Lett.*, **25**, 149 (1970).
33. D. Chandler, J. D. Weeks, and H. C. Andersen, *J. Chem. Phys.*, **54**, 5237 (1971).
34. H. C. Andersen, D. Chandler, and J. D. Weeks, *Phys. Rev. A*, **4**, 1597 (1971).
35. B. B. Laird, *J. Chem. Phys.*, **115**, 2887 (2001).
36. R. L. Davidchack and B. B. Laird, *J. Chem. Phys.*, **118**, 7651 (2003).
37. J. Q. Broughton and G. H. Gilmer, *J. Chem. Phys.*, **84**, 5741 (1986).
38. M. Heymann and E. Vanden-Eijnden, *Phys. Rev. Lett.*, **100**, 140601 (2008).

EFFECTS OF PROTEIN SIZE ON THE HIGH-CONCENTRATION/LOW-CONCENTRATION PHASE TRANSITION

PATRICK GROSFILS

Center for Nonlinear Phenomena and Complex Systems CP 231,
Université Libre de Bruxelles,
Blvd. du Triomphe, 1050 Brussels, Belgium

CONTENTS

I. INTRODUCTION

Proteins in solution show a variety of phase transitions, conformational transitions, and aggregation phenomena that are the subjects of active research in different areas. Understanding protein condensation is, for example, important in biomedical research because certain diseases, such as human cataracts, result from this process [1]. Protein crystallization is another subject of extensive research as protein

Advances in Chemical Physics, Volume 151: Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanoscale Materials, First Edition. Edited by Gregoire Nicolis and Dominique Maes.

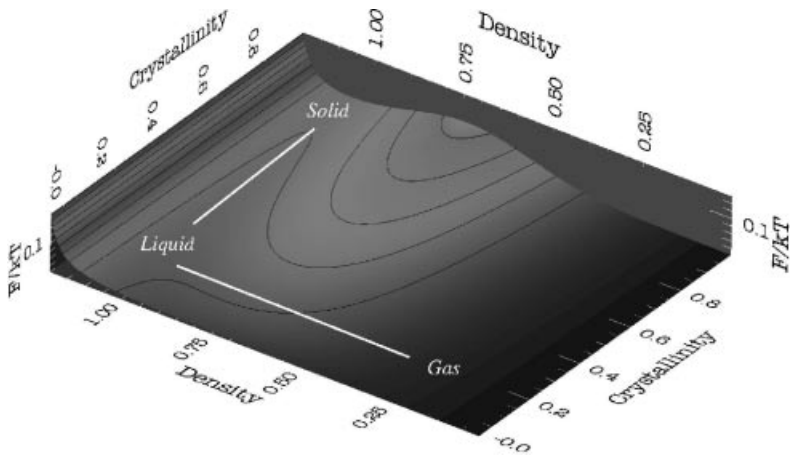


Figure 1. Two-step crystallization pathway from proteins in solution (gas) to a crystal (solid) via an intermediate metastable dense cluster (liquid).

crystals of good quality are required in X-ray crystallography for protein structure determination. A large body of work has been devoted to this topic and especially to nucleation, the process that initiates protein condensation.

It follows from various numerical, theoretical, and experimental studies that protein nucleation does not always conform to the classical nucleation theory, but it is often subject to a more complex process. According to the proposed new theory, protein crystallization is a two-step process that proceeds via an intermediate metastable protein-rich state similar to the liquid state [2–6]. Thus, inside the solution, a dense cluster is first formed, followed by a rearrangement of the molecules within the cluster into an ordered structure (see Fig. 1). This mechanism appears to apply to many crystallization processes from solutions irrespective of structure and size of molecules. One way in which the molecular size can affect nucleation is through its impact on stability of the dense phase. Indeed, for potentials with a range sufficiently short with respect to the excluded volume, the dilute-phase/dense-phase transition becomes metastable with respect to the dilute-phase/crystal transition [6]. Moreover, the size of the molecule appears to have also an impact on the second step of the nucleation process, that is, protein rich to crystal transition, through an increase in organization time that goes with a greater molecular size [7].

In this chapter, the effects of the molecular size on the metastable liquid phase are examined in the light of results obtained by simulating a model of globular protein using the Monte Carlo method. Particular attention is paid to interfacial properties that have so far been little studied. The coexistence curve and the interfacial tension

are determined for different molecular sizes. Globular proteins constitute one of the main classes of proteins. The spherical structure is here induced by their tertiary structure that results from the hydrophobic effect. In the next section, it is shown how these biophysical informations can be incorporated into a model potential. The Monte Carlo simulation method is then described and some problems inherent to metastable phase and steep interaction potential are discussed. This is followed by a section presenting and discussing the results of the simulations. In the concluding section, we discuss in the light of our finding a possible explanation of why in some systems intermediate structured units are first built before their incorporation to a growing phase.

II. MODELING STRUCTURE AND INTERACTION OF GLOBULAR PROTEINS

Proteins are long-chain molecules with specific monomer sequences that fold and form compact and ordered arrangements. One view of the protein folding process assumes that it is the nonlocal interaction between amino acids that are widely separated in the sequence that drives the collapse of the chain and gives the protein its structure. The idea, coming from polymer modeling, is that protein chain contains two kinds of residues that interact strongly with water and contribute predominantly to the folding process [8, 9]. Nonpolar segments interact unfavorably with water and during folding they tend to be buried in the protein interior away from contact with water. Polar segments, on the contrary, interact favorably and very strongly with water and form the outer shell of the folded protein. This hydrophobic effect is the major force for folding proteins and stabilizing their structure [10].

Most of the theoretical models that have been introduced to study protein crystallization or phase coexistence are based on the idea that under conditions that favor native state, proteins can be approximated as colloids with short-range attractions. This suggests the use of an effective potential for the theoretical study of globular proteins in solution. Since this model ignores the polymeric character of the protein, it might seem that results drawn from this approach could hardly be generalized to real protein solutions. Actually, information such as the amino acid sequence can be integrated into an effective protein–protein interaction potential using protein folding models. Of particular interest is the pair potential

$$V(r) = \frac{4\epsilon}{\alpha^2}\left(\frac{1}{((r/\sigma)^2-1)^6} - \frac{\alpha}{((r/\sigma)^2-1)^3}\right) \tag{1}$$

originally proposed by ten Wolde and Frenkel [3] to study the phase behavior of globular proteins in solution. As briefly discussed next, the parameters entering in Eq. (1) can be related to average structural characteristics of native and denatured protein states using Dill's heteropolymer collapse theory [8, 11].

According to Dill's theory, proteins are heteropolymers whose segments can be considered as hydrophobic or polar. Such a two-species model is in contrast to the biophysical representation of the primary protein structure in terms of 20 amino acids. This seemingly oversimplified description is nevertheless able to emulate the collapse of the chain driven by the formation of hydrophobic contacts and a loss of conformational entropy. Also, the state of the protein, native or denatured, is deduced depending on sequence and the temperature. The model also predicts that in the native state, there are less hydrophobic segments exposed to the solvent than in the denatured state.

Heteropolymer collapse (HPC) theory is based on the principle that interactions between hydrophobic segments and the solvent are less favorable than those between polar segments and the solvent. The theory integrates this principle by means of Kauzmann's model [12] that treats the protein interior as a liquid so that the hydrophobic interaction can be evaluated as the change of free energy that results from the transfer of a hydrophobic side chain from the solvent (H_2O) to the interior of the protein considered as a pure hydrophobic liquid [13–17]. The nature of each segment, that is, hydrophobic or polar, is taken into account statistically by Φ, the fraction of segments that can be considered as hydrophobic among the N_s segments that comprise the chain. From this information, the theory predicts the size of the protein, expressed by a radius of gyration, in the native (R_{native}) and denatured ($R_{\text{denatured}}$) states. Because these structural characteristics of native and denatured proteins follow from the sequence of hydrophobic/polar segments along the chain, they are affected by Φ in such a way that increasing the hydrophobicity will lead to a more compact protein state.

Interaction between proteins follows from the same principle that governs protein folding, namely, the hydrophobic effect. HPC theory predicts the fraction of segments in contact with the solvent (f_e) and $\theta(T)$, the fraction of hydrophobic residues present on the surface. From $\theta(T)$, the protein–protein interaction is derived if one assumes that protein interaction follows from the desire to desolvate the hydrophobic surface segments by burying them into hydrophobic patches on interacting proteins.

In short, protein structure, stability, and interaction are closely connected through the primary structure of the protein, that is, the amino acid sequence, with the result that interaction strength should depend on protein state. This is evidenced by the fact that in the denatured state, the protein is larger than in its native state and its surface contains more hydrophobic segments, and as a result the interaction energy becomes state dependent. All this information is incorporated into the effective potential (1) by allowing different interaction energies, ϵ_{NN}, ϵ_{DD}, or ϵ_{ND}, and different contact distances, σ_{NN}, σ_{DD}, or σ_{DN}, for interaction between native states (NN), denatured states (DD), or native and denatured states (ND).

Protein hydrophobicity is an important parameter that drives protein stability, because highly hydrophobic proteins destabilize more easily through the formation

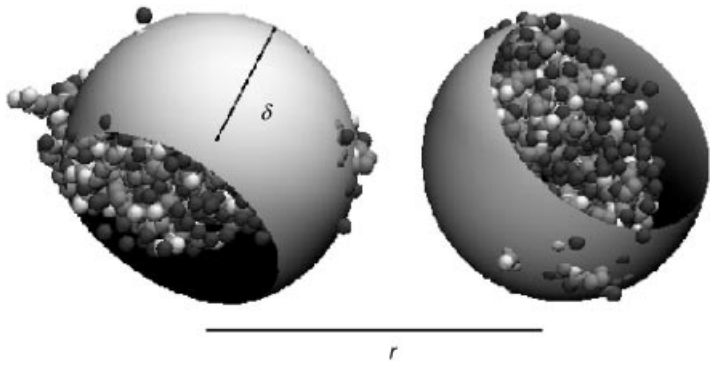


Figure 2. Coarse-grained model for protein structure and protein interaction. The native state of the protein becomes a spherical shell with a radius δ that depends on the number of amino acids. The protein–protein interaction (2) is a function of the distance r between the centers of the shells.

of contacts between proteins. This aspect of the problem is not explored in this chapter, where only the effects of the size parameter on phase behavior are analyzed. It would, however, be interesting to investigate the effect of a temperature-dependent strength on the phase diagram and to take into account structural fluctuations of the protein state that have an impact on the protein solubility. While such an effect could, in principle, be incorporated by using a detailed protein model, this is for the moment much too complicated. Thus, a coarse-grained model derived from HPC theory as represented in Fig. 2 is a good compromise as it incorporates some effects of the protein structure while being at the same time computationally efficient.

III. METHODOLOGY

A. Model and Monte Carlo Method

The behavior of protein solutions as a function of the molecule size and potential range has been investigated by Monte Carlo simulations. The model potential used is based on Eq. (1) slightly generalized to include the molecule size δ,

$$V(r) = \frac{4\epsilon}{\alpha^2}\left(\frac{1}{(r_\star^2 - \delta^2)^6} - \frac{\alpha}{(r_\star^2 - \delta^2)^3}\right) \tag{2}$$

where $r_\star$ is the reduced distance r/σ. As the ten Wolde–Frenkel original potential, the potential in Eq. (2) captures the essential properties of protein solutions and in particular the fact that for sufficiently short-range attractions, the liquid–liquid coexistence becomes thermodynamically metastable. According to generally accepted criteria, this should occur when the range of the attraction is ∼25% of the

range of the repulsive part that for Eq. (1) is expected to be found when $\alpha \simeq 10$ [18]. Here, we consider Eq. (2) starting with $\alpha = 1$ and $\delta = 0$, which corresponds to the Lennard-Jones potential, and increase the molecular size δ at different values α in order to study the transition between a simple fluid that possesses a stable liquid state to a fluid characterized by a metastable liquid–liquid coexistence curve. Because the parameters entering the potential are fixed during each Monte Carlo simulation, configurational fluctuations of the protein, such as native–denatured transition, are not incorporated into this analysis.

To characterize the metastable phase, we measured the coexistence densities and the surface tension at different temperatures using a standard Metropolis Monte Carlo algorithm (MC-NVT) [19]. This method raised some problems because at low temperature the dense fluid tends to decay to the stable crystal phase. This is the reason why the Gibbs ensemble Monte Carlo [20] or the grand canonical Monte Carlo [19] methods are usually preferred to measure metastable coexistence curve.

A number of procedures exist that prevent crystallization of the liquid in a simulation. One possibility is to use the restricted Monte Carlo method [21] or to impose some geometric constraints on the molecular configurations, such as the number of neighbors per particle. Our solution to this problem was to start from a Lennard-Jones fluid ($\delta = 0$ and $\alpha = 1$) in the stable two-phase region and to gradually modify δ until reaching the desired molecular size. Then the system was simulated in a temperature range starting from the high temperature to the lowest achievable temperature, that is, the temperature at which we could not prevent crystallization.

B. Surface Tension Measurement

Many surface tension calculation methods have been developed over the years. Clearly, the best method is the one that gives a good estimate of the surface tension, that is, with small numerical errors, at the lowest computational cost. In the present case, however, the selection has proven to be difficult (see discussion below) and different approaches to measure the surface tension were tested. Most of the approaches implemented had only limited success for reasons relating to the range of the effective potential or to the instability of the liquid phase.

In the mechanical approach, the surface tension is derived from a microscopic definition of the pressure tensor [22, 23]. This is a widely used method that requires computing the virial that in the case of steep and short-range potential can cause numerical difficulties.

Thermodynamic approaches allow to calculate the surface tension without evaluating the virial but by perturbating the system by increasing or decreasing the interface area by a small amount. The way this is implemented is in some cases problematic. To understand the cause of the problem, consider the Bennett method

[24], where the surface tension γ is obtained by measuring the relative free energy between two systems differing only by their interface areas or more explicitly using the definition

$$\gamma = \left(\frac{\partial F}{\partial A}\right)_{N,V,T} \tag{3}$$

where F is the free energy and A is the area of the interface. In practice, the energies of two configurations of the systems with equal volumes but with interface areas $A_0 \neq A_1$ should be compared. Because the surface change is done at constant volume, the system is expanded along one direction and compressed along the transverse direction. The difficulty lies in the choice of the magnitude of the expansion–compression that should give a reliable average of the ensuing free energy change. In our case, the compression along one direction generates most of the time particle configurations with many particle overlaps due mainly to the size of the excluded volume with respect to the small attraction range of the potential. Besides, the compression is also a major source of difficulty for the stability of the liquid phase because compression may trigger crystallization of the liquid slab. One solution is to vary very slightly the interface, which implies small free energy differences. Unfortunately, it turns out that the surface tension of this system is particularly small, so large perturbations are necessary to have a measurable free energy difference.

We have therefore used instead the interface wandering approach [25]. This method relies on an analysis of the interface area distribution function that is allowed to fluctuate when the system is subject to shape perturbations, which means that the positions of the particles are rescaled so as to preserve the volume of the system [26]: $x' = x\,(A_1/A_0)^{1/2}$, $y' = y\,(A_1/A_0)^{1/2}$, and $z' = z\,(A_0/A_1)$. In practice, the system is perturbed according to this scaling procedure and the rescaled configuration is allowed to replace the previous configuration. This is in contrast to what is done in the Bennett method, where the rescaled configuration does not replace the previous configuration but is used to obtain only the energy difference between configurations with different interface areas. As a result of configuration change, the interface area wanders driven by competing pressures, $p_{\parallel}$ and $p_{\perp}$, respectively, parallel and perpendicular to the interface [25]. Consequently, the interface area A becomes distributed according to

$$f(A) = e^{-\beta\,(2\,\gamma A - p\,V)} \tag{4}$$

where γ is surface tension, V is the volume, and $\beta = 1/k_{\mathrm{B}}T$. If interface area is collected during simulation, a fit of Eq. (4) through the observed distribution of areas gives the surface tension value. According to Eq. (4), for positive surface tension, the most probable interface area is the smallest one that in the absence of any competitive constraint leads to a vanishing interface. This problem can

be solved either by imposing additional work on the system to thwart the effects of surface tension or by limiting the fluctuations of the area within an interval $[A_{\min}, A_{\max}]$. The latter method was chosen to measure the area distribution law in our simulations. Practically, an area A_1 is chosen randomly within the interval and this new value is accepted with the probability

$$P_{\text{accepted}} = \min\left(1,\ e^{-\beta(E_1-E_0)}\right) \tag{5}$$

where E_0 is the previous energy of the system, with an area A_0, and E_1 is the energy obtained after rescaling to have an area A_1. The method requires the use of very small changes in the shape to avoid particles overlaps. As shown Fig. 3, the numerical value of the surface tension can nevertheless be extracted from the distribution law despite the random fluctuations inherent to small systems.

As regards technical details of the simulations, the systems simulated consisted of $N = 1885$ molecules contained in a region of fixed volume V kept at constant temperature T. The interaction potential between two molecules separated by a distance r is Eq. (2) with a cutoff at $r_c = 2.8$ and the potential is not shifted. The simulation box with dimensions $L_x = L_y = 9\sigma$ and $L_z = 108\sigma$ contained a slab of proteins in the liquid phase surrounded along the z-direction by two slabs of proteins in the gas phase. The thickness of the liquid slab ($\Delta z \simeq 27\sigma$) was such that most of the proteins were in the liquid phase. Periodic boundary

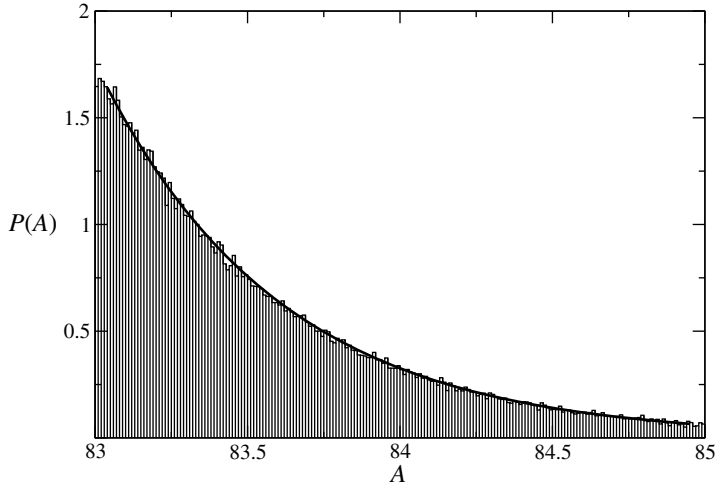


Figure 3. The distribution function of interface area $P(A)$ as determined by the interface wandering method. The distribution law given is fitted by $f(A) \sim e^{-2\beta\gamma}$, which gives a precise estimate of the surface tension γ.

conditions were imposed on the system in the three directions. The systems were equilibrated during 5×10^5 Monte Carlo updates per molecule. To ensure the stability of the liquid phase, the equilibration step started from a stable molecular configuration obtained in a previous simulation at $(T + \Delta T, \alpha + \Delta\alpha, \delta + \Delta\delta)$ and slowly modified these parameters until reaching (T, α, δ). Then during 10^6 Monte Carlo cycles, densities were measured and box rescalings were performed to get interface areas distributed between $A_{\min}$ and $A_{\max}$.

IV. MOLECULAR SIZE EFFECTS ON THE METASTABLE PROTEIN-RICH/PROTEIN-POOR PHASE COEXISTENCE

The molecular size impacts on the metastable liquid phase through an excluded volume effect. This effect can be observed particularly on the phase coexistence curve and on the interfacial tension in the metastable region of the phase diagram. Globular proteins are modeled typically using $\alpha = 50$ and $\delta = 1$. However, simulations were mostly performed for $\alpha = 1$ and 5 at excluded volume parameter δ between 0 and 1, the reason being that at $\alpha = 50$ the interparticle potential is very steep and the coexistence curve lies deep inside the metastable region, which, as discussed above, makes simulation extremely difficult.

Coexistence curves for different protein sizes obtained for the cases $\alpha = 1$ and $\alpha = 5$ are shown in Fig. 4. The simulation data are displayed together with theoretical curves as predicted by the theory [27, 28]. The phase diagram pertaining to globular proteins, which corresponds to $\alpha = 50$ and $\delta = 1$, is shown in Fig. 5. The figure contains, in addition to the data obtained in this work, some original data obtained by ten Wolde and Frenkel using the Gibbs ensemble Monte Carlo method and therefore correspond to a physically different situation, as in this case there is no interface between the liquid and the gas phase [3].

By merely looking at these figures, one can easily see the dramatic effect of the molecular size on the phase diagram. Interestingly, the phase diagram is hardly affected by the change in the potential range α for small excluded volume ($\delta < 0.4$).

An estimate of the critical density and temperature for each system is given in Table I. The data clearly indicate that an increase in the molecular size results in lowering of the critical temperature and that this effect becomes more pronounced with increasing α. However, there are similarities in the thermodynamic properties of these systems as evidenced by Fig. 6, which shows coexistence curves with the density and the temperature scaled to the estimated critical density and critical temperature, respectively, although it is apparent that systems with $\delta > 0.6$ tend to deviate from scaling. This is presumably due in part to a larger error in the estimation of the critical parameters, but this could also indicate that, in addition to the temperature and density, a third parameter related to the volume of the particle must be included in the law of corresponding states.

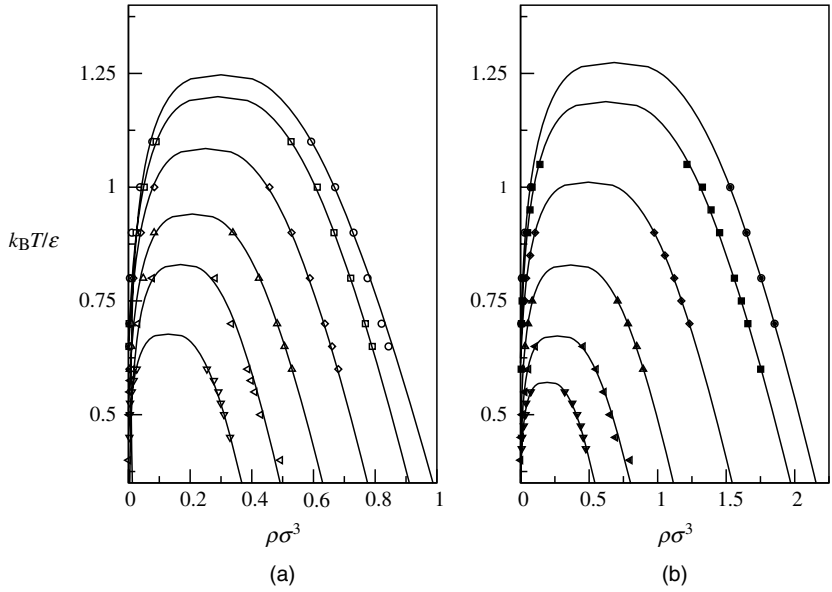


Figure 4. The phase diagrams for $\alpha = 1$ (**a**) and for $\alpha = 5$ (**b**), as determined by simulation and theory for different values of the molecular size δ. At fixed temperature, the points on the left part of the curves give the protein concentration in the protein-poor phase and points on the right part of the curve give the protein concentration in protein-rich phase. The curves, from top to bottom, are for reduced protein size $\delta = 0,\ 0.2, 0.4, 0.6, 0.8$ and 1.0, respectively.

There is a relation, noticed by Vliegenhart and Lekkerkerker, that relates the critical temperature of many different systems to the volume of their particles. In Ref. 29, these authors showed that the second virial coefficient

$$B_2(T) = 2\pi \int_0^\infty \left(1 - \exp\left(-\frac{V(r)}{k_B\, T}\right)\right) r^2\, dr \tag{6}$$

when evaluated at the critical temperature T_c appears to lie in a narrow range around $-6\, v_0$, where $v_0 = \pi/6\, r_0^3$ is the volume and the particle diameter r_0 is the distance at which the potential crosses zero. In the case of potential (2), $r_0/\sigma = \sqrt{(1/\alpha)^{1/3} + \delta^2}$.

The relation between B_2 and v_0 has been tested by evaluating this coefficient at the critical temperatures obtained from the coexistence curves. The results of these calculations are listed in Table II. We note that the values of $B_2(T_c)/v_0$ are close to -6, but this value tends to decrease when the volume of the molecule increases. This behavior seems to be confirmed by experimental data as evidenced by the

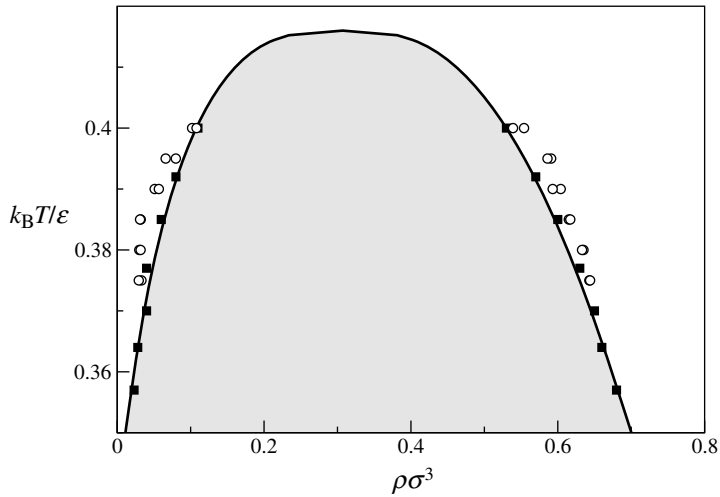


Figure 5. The phase diagram for the model of globular protein ($\alpha = 50$ and $\delta = 1.0$). At fixed temperature, protein concentrations located inside the shaded region separate into protein-rich and protein-poor domains inside which the protein concentration is given by the two points on the coexistence curve. The figure includes the theoretical curve and the simulation data, shown as circles (this work) and squares (from Ref. 3).

analysis of the experimental data reported in Ref. 29 and that concern proteins of widely differing sizes.

More interesting is the possibility to predict the critical temperature from the relationship

$$B_2(T_c) = -6\, v_0 \tag{7}$$

TABLE I
Critical Properties Estimated from the Coexistence Data for $\alpha = 1$ and $\alpha = 5$

	$\alpha = 1$		$\alpha = 5$	
δ	T_c (coex)	ρ_c (coex)	T_c (coex)	ρ_c (coex)
0.0	1.24	0.31	1.28	0.68
0.2	1.19	0.29	1.20	0.62
0.4	1.08	0.25	1.01	0.49
0.6	0.93	0.21	0.83	0.37
0.8	0.80	0.17	0.66	0.27
1.0	0.66	0.15	0.57	0.19

Adapted from Ref. 31.

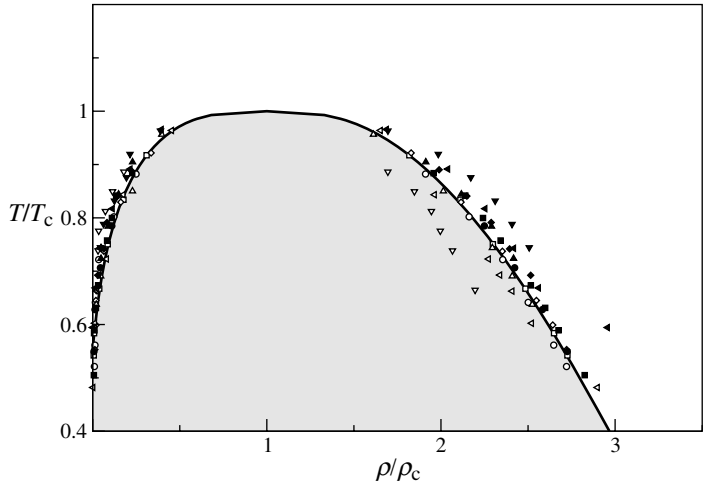


Figure 6. As in Figs. 4 and 5 with the density and temperature scaled to the critical density and temperature for each potential.

TABLE II
Second Virial Coefficient B_2/v_0 at $T = T_c$ for $\alpha = 1$ and $\alpha = 5$

δ	$\alpha = 1$	$\alpha = 5$
0.0	−6.33	−6.26
0.2	−6.36	−6.27
0.4	−6.38	−6.58
0.6	−6.70	−6.86
0.8	−6.93	−7.20
1.0	−7.87	−7.62

Taking it for granted, we used it to compute the critical temperature T_c. In Table III, the critical temperatures computed in this way are compared to the values obtained from the coexistence curves. It shows that critical temperatures T_c given by the empirical relationship approach the values of T_c from the simulations to within $\sim$2% for $\delta < 0.6$ and is precise to within 10% for $\delta \geq 0.6$, with each time a computed temperature greater than the one measured. Once again the critical temperature is seen to depend on the molecular size, but in a more pronounced manner than expected from Eq. (7).

TABLE III
Critical Temperature Estimated from Coexistence Compared to the Criteria $B_2(T_c) = -6\,v_0$, for $\alpha = 1$ and $\alpha = 5$

	$\alpha = 1$		$\alpha = 5$	
δ	T_c (coex)	$T_c(B_2 = -6v_0)$	T_c (coex)	$T_c(B_2 = -6v_0)$
0.0	1.24	1.27	1.28	1.31
0.2	1.19	1.22	1.20	1.22
0.4	1.08	1.11	1.01	1.06
0.6	0.93	0.98	0.83	0.88
0.8	0.80	0.86	0.66	0.75
1.0	0.66	0.75	0.57	0.64

V. SURFACE TENSION

The results for the surface tension as a function of δ are shown in Fig. 7 for $\alpha = 1$ and $\alpha = 5$. The figure also shows a representation of the coarse-grained protein model with a scaled excluded volume. The data clearly show a lowering of the surface tension when increasing the size of the molecule at fixed α. Furthermore,

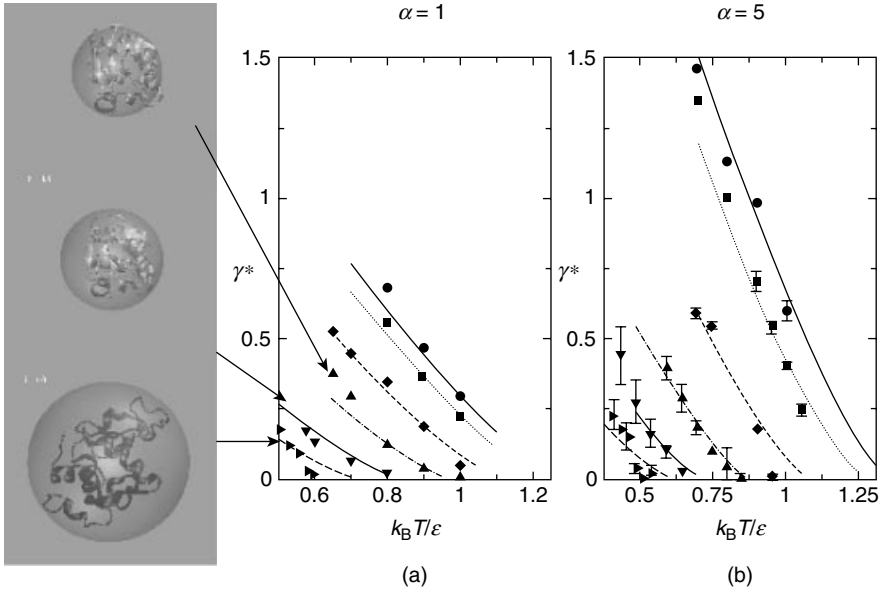


Figure 7. Surface tension as a function of the temperature for $\alpha = 1$ (**a**) and $\alpha = 5$ (**b**) determined by Monte Carlo simulations and theory [31].

TABLE IV
Fits to Surface Tension: $\gamma(T) = \gamma_0 T_c (1 - T/T_c)^{1.26}$

	$\alpha = 1$		$\alpha = 5$	
δ	T_c	γ_0	T_c	γ_0
0.0	1.21	2.15	1.33	2.83
0.2	1.18	1.95	1.18	3.53
0.4	1.09	1.55	0.99	2.96
0.6	0.92	1.87	0.85	2.09
0.8	0.80	1.68	0.66	2.63
1.0	0.61	2.52	0.54	3.06

for $\delta > 0.4$, the variation of the surface tension with the attraction range, fixed by α, is very low. Assuming that these systems obey a simple corresponding states law with only two parameters, the scaled surface tension (with unit molar mass)

$$\gamma_s = \frac{\gamma}{T_c \rho_c^{2/3}} \tag{8}$$

obeys the relation [30]

$$\gamma_s = \gamma_0 (1 - T/T_c)^{1.26} \tag{9}$$

where γ_0 is assumed to be a constant. This relation is known to describe well the surface tension of simple fluids and in particular fluids with spherical symmetric potentials as the Lennard-Jones potential [28].

The results of a fit of the surface tensions to the form $\gamma = \gamma_0 T_c (1 - T/T_c)^{1.26}$ are shown in Table IV for $\alpha = 1$ and 5 and in Table V for $\alpha = 50$, and the scaled surface tensions are shown in Fig. 8a as functions of $1 - T/T_c$. While the relation (9) is relatively well respected for surface tensions evaluated at fixed α, the

TABLE V
Surface Tension $\gamma(T)$ for $\alpha = 50$ and $\delta = 1.0$

	$\alpha = 50$
T	γ
0.375	0.1199
0.385	0.0803
0.390	0.0567
0.395	0.0350
0.400	3.8×10^{-5}

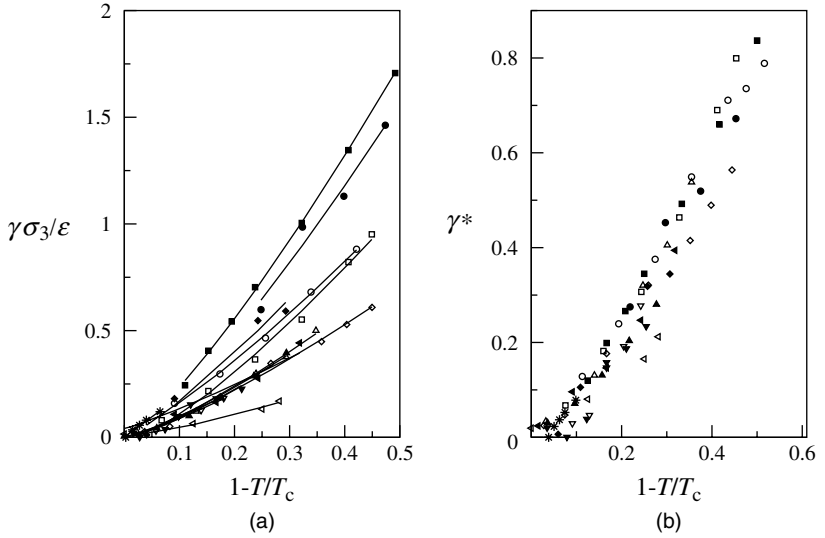


Figure 8. (**a**) Surface tension γ as a function of the distance from the critical point. (**b**) Scaled surface tension $\gamma^\star$ for all different values of α and for different molecular size δ. The scaled surface tension shows a behavior that obeys the extended corresponding states law.

curves for different α values do not superposed as may be expected from a simple application of the law of corresponding states.

The question then arises as to the possibility to extend the law of corresponding relation by including, in an appropriate way, the parameter α. To this end, we use the fact that the surface tension scales with the critical temperature as can be deduced from the fact that γ_0 is almost constant. The critical temperature, on the other hand, is affected by δ, which fixes the ratio between the repulsion range, $\sqrt{(2/\alpha)^{1/3} + \delta^2}$, and the attractive range of the potential. So, as argued in the previous section, the critical temperature decreases with increasing δ, but changing α when $\delta = 1$ has relatively little effect compared to the effect obtained when $\delta = 0$. In the latter case, one finds a repulsion with a range that scales as $\alpha^{-1/6}$ so that one expects a surface tension that varies as $\gamma = \alpha^{1/3}$. In other words, the surface tension depends on the number of particles that a fixed surface area may contain. To verify the previous reasoning, a scaled surface tension, $\gamma^\star = \gamma/T_c\sigma^{2/3}\alpha^{1/3}$, is represented as a function of the distance from the critical temperature $1 - T/T_c$. The result is shown in Fig 8b for $\alpha = 1$ and 5 for all values of δ and for $\alpha = 50$ with $\delta = 1$. It shows that despite a wide range of α and δ values, the scaled surface tension obeys to a good approximation this extension to the law of corresponding states.

VI. NUCLEATION AND PROTEIN SIZE

In the process of protein crystallization from solution, the metastable liquid phase is thought to play a key role as it is within it that a crystalline nucleus forms [2, 3, 6]. According to the classical nucleation theory, the formation of a droplet of radius R has excess free energy

$$\Delta\Omega = \frac{4\pi}{3} R^3 \left(\omega(\rho_l) - \omega(\rho_v)\right) + 4\pi R^2 \gamma \tag{10}$$

where $\omega(\rho) = f(\rho) - \mu\rho$, $f(\rho)$ is the bulk Helmholtz free energy per unit volume of the fluid at density ρ, μ is the chemical potential, and γ is the dilute-phase/dense-phase surface tension at coexistence. The first term is the free energy change due to the phase transition, while the last term is the free energy change requires for the formation of the surface. In the transition from protein-poor to protein-rich phases, the first term describes the spontaneous tendency to form a protein-rich droplet that as the protein-rich phase is more stable than the protein-poor phase is a negative term. The interface term, on the other hand, increases the free energy by an amount proportional to the surface area of the liquid droplet. As a result, there is a critical radius, $R_c = 2\gamma/\Delta\omega$, for which the excess free energy is the greatest and whose value at $R = R_c$, $\Delta\Omega_{max} = 16\pi\gamma^3/3\Delta\omega$, defines the barrier for nucleation.

On the basis of the above discussion concerning the relation between surface tension and protein size, one can conclude that the free energy barrier will decrease with increasing molecular size δ and it will be affected to a lesser extent by the range of the potential α. These results suggest that increasing, even slightly, the effective excluded volume of a molecule will lead to a decrease in the surface tension and hence decrease the nucleation barrier. The question then is whether the excluded volume can be modified experimentally with the objective to control the nucleation process. According to the HPC theory presented in Section II, the protein diameter depends on the state of the protein, native or denatured, and therefore it must be possible to control its value by modifying the hydrophobic interactions through a change in pH. An important aspect of the transition between native and denatured states is that it is a phenomenon that occurs spontaneously through thermal fluctuations, and therefore molecular size variations do exist at all temperatures.

Another important aspect of the link between surface tension and molecular size is the support it gives to some growth mechanisms and to the fact that in some cases intermediate structures are formed prior to incorporation into a growing cluster. A typical example is the large population of tetrameric structure often observed during droplet formation and, presumably, during crystallization as well. If one assumes that a tetramer is nothing more than a big molecule whose volume is four times the volume of a monomer but keeping the same interaction range, then the

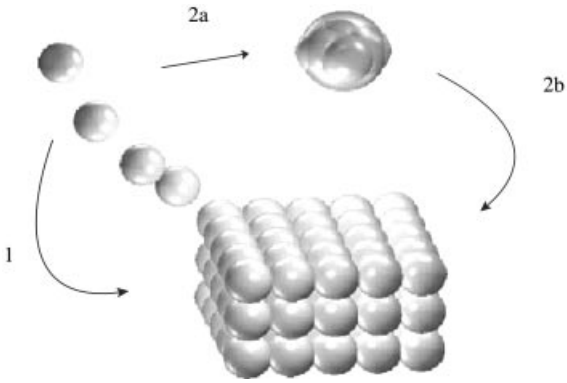


Figure 9. The two pathways for particles aggregation. In the first pathway (1) the monomers are directly incorporated to the cluster. The second pathway is a two-step mechanism that contains as intermediate step the formation of a tetramer (2a), which is then incorporated into the cluster (2b).

surface tension of the tetrameric system will be smaller than the system made of monomers. As shown in Fig. 9, instead of the aggregation pathway that consists of adding monomers, there is an alternative pathway, energetically more efficient and that, as an intermediate step, goes through the formation of tetramers. The explanation given here for a preferred aggregation pathway should, of course, be supported by theoretical work. However, the important advantage of the proposed mechanism is that it is suitable for further theoretical analysis.

VII. CONCLUSION AND FINAL REMARKS

In this chapter, the effects of molecular size on the metastable dilute liquid–dense liquid phase coexistence have been discussed on the basis of results obtained by the Monte Carlo method. The behavior of the coexistence curve and the protein-rich/protein-poor surface tension have been analyzed as a function of the molecular size. The starting point of the method is the construction of a pair potential that allows to move smoothly from a simple fluid, that is, Lennard-Jones interaction, to the short-ranged, "hard Lennard-Jones" used as a model for globular protein interactions. The potential contains as parameters the molecular size, δ, and the range of the potential, controlled by the dimensionless parameter α. It has been briefly outlined how in the framework of HPC theory these parameters can be related to biophysical characteristics of proteins in solution.

The key result of this study is that it is the increase in the hard-core diameter rather than the decrease in range of the potential that causes a dramatic decrease in the surface tension and has the most important impact on the protein-rich/protein-poor phase coexistence. However, despite the changes arising from these modifications, the validity of the law of corresponding states can be preserved when extended by introducing another length scale as a new parameter.

An important aspect of the relation between surface tension and molecular size is its implication for the nucleation process. Thus, all other things being equal, an increase in the molecular size leads to a decrease in the nucleation barrier. From this result, we suggest that the apparition of small aggregates during the growing phase of droplet or crystal is a necessary intermediate step because it helps the nucleation process through a decrease in the nucleation barrier. This two-steps mechanism should also apply to other systems that present intermediate structures that self-assemble into superstructures such as zeolites.

There are many ways in which this work can be improved. First, the model potential can be refined to capture more specific properties of protein systems. One development of this kind currently receiving attention is anisotropic attraction [1]. Patchy interactions better capture the inhomogeneous surface of globular proteins. Another direction for refinement is to take into account the nonrigidity of the protein and the fluctuations between native and denatured states. In the denatured state, the number of hydrophobic residues at the surface is larger than the number in the native state. Therefore, denatured state aggregates more easily because the hydrophobic attraction is stronger. Therefore, including transitions between protein states could significantly modify the behavior of protein solutions, but in a way that does not contradict the conclusions of this chapter.

Acknowledgments

I would like to thank J. F. Lutsko for discussions and for his insightful comments on the chapter. This work was supported by the European Space Agency under contract number ESA AO-2004-070.

References

1. J.D. Gunton, A. Shiryayev, and D.L. Pagan, *Protein Condensation: Kinetic Pathways to Crystallization and Disease*, Cambridge University Press, 2007.
2. P. Vekilov, *Cryst. Growth Des.*, **4**, 671 (2004).
3. ten Wolde and Frenkel, *Science*, **77**, 1975 (1997).
4. G. Nicolis and C. Nicolis, *Physica A*, **351**, 22 (2005).
5. A. Shiryayev and J. D. Gunton, *J. Chem. Phys.*, **120**, 8318 (2004).
6. J. F. Lutsko and G. Nicolis, *J. Chem. Phys.*, **122**, 244907 (2005).
7. D. Erdemir, A. Y. Lee, and A. S. Myerson, *Acc. Chem. Res.*, **42**, 621 (2009).
8. K. A. Dill, *Biochemistry*, **24**, 1501 (1985).
9. K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, *Protein Sci.*, **4**, 561 (1995).
10. K. A. Dill, *Protein Sci.*, **8**, 1166 (1999).
11. J. K. Cheung and T. M. Truskett, *Biophys. J.*, **89**, 2372 (2005).
12. W. Kauzmann, *Adv. Protein Chem.*, **14**, 1 (1959).
13. C. Tanford, *The hydrophobic Effect*, Wiley, New York, 1980.
14. C. Tanford, *J. Am. Chem. Soc.*, **84**, 4240, (1962).
15. C. Tanford, *J. Am. Chem. Soc.*, **86**, 2050, (1964).

16. Y. Nozaki and C. Tanford, *J. Biol. Chem.*, **246**, 2211, (1971).
17. D. Chandler, *Nature*, **437**, 640 (2005).
18. J. F. Lutsko and G. Nicolis, *Phys. Rev. Lett.*, **96**, 046102 (2006).
19. D. Frenkel and B. Smit, *Understanding Molecular Simulation*, Academic Press, Inc., Orlando, FL, 2001.
20. A. Z. Panagiotopoulos, *Mol. Phys.*, **61**, 813 (1987).
21. D. S. Corti and P. G. Debenedetti, *Chem. Eng. Sci.*, **49**, 2717 (1994).
22. J. G. Kirkwood and F. P. Buff, *J. Chem. Phys.*, **17**, 338 (1949).
23. J. H. Irving and J. G. Kirkwood, *J. Chem. Phys.*, **18**, 817 (1950).
24. C. H. Bennett, *J. Comput. Phys.*, **22**, 245 (1976).
25. L. G. MacDowell and P. Bryk, *Phys. Rev. E*, **75**, 061609 (2007).
26. E. Salomon and M. Mareschal, *J. Phys. Condens. Matter*, **3**, 3645 (1991).
27. N. B. Wilding, *Phys. Rev. E.*, **52**, 602 (1995).
28. P. Grosfils and J. F. Lutsko, *J. Chem. Phys.*, **130**, 054703 (2009).
29. G. A. Vliegenthart and H. N. W. Lekkerkerker, *J. Chem. Phys.*, **112**, 5364 (2000).
30. E. A. Guggenheim, *J. Chem. Phys.*, **13**, 253 (1945).
31. P. Grosfils and J. F. Lutsko, *Langmuir*, **26**, 8510 (2010).

GEOMETRIC CONSTRAINTS IN THE SELF-ASSEMBLY OF MINERAL DENDRITES AND PLATELETS

JOHN J. KOZAK

DePaul University, 243 South Wabash Avenue, Chicago, IL, USA

CONTENTS

I. INTRODUCTION

Models of platelet and dendrite formation have been developed and studied extensively in the literature for generations [1]. Insights drawn from extensive experimental and theoretical studies on a related problem, thin film growth, have done much to clarify the factors determining specific island morphologies. As elaborated by Lagally and coworkers [2], there are, broadly speaking, two classes of shapes, one compact with relatively straight and equiaxed island edges and the other fractal-like having rough island edges or highly anisotropic shapes. In this chapter, we focus on geometrical factors governing the self-assembly of island morphologies by considering different scenarios for the sequential deposition of adatoms.

Advances in Chemical Physics, Volume 151: Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanoscale Materials, First Edition. Edited by Gregoire Nicolis and Dominique Maes.

The patterns displayed by inorganic dendrites and platelets are among Nature's most remarkable, long celebrated using the imagery of language but first described in terms of a "formative principle" by Kepler. Four centuries ago, his 1611 treatise, "On the Six-Cornered Snowflake," launched the field of "self-assembly," with advances since then reflected in the diverse contributions presented at the Brussels Workshop. In the concluding section of this chapter, we discuss the relevance of our results to dendrite and platelet formation in two mineral systems, the hollandite group of manganese oxides and zincite (zinc oxide).

To account for excluded volume effects and attendant steric constraints influencing the immobilization of a randomly diffusing adatom, a lattice model is introduced and the efficiency of "docking" at a particular site j is studied using the theory of finite Markov processes. By solving the underlying stochastic master equation, the (time-dependent) probability distribution function $\rho(t)$ describing the process can be determined. In this chapter, we shall report values of the first moment of $\rho(t)$, the average number $\langle n \rangle_j$ of displacements of the diffusing particle before being trapped at site j, a useful, intuitive measure of the efficiency of the underlying process.

Once $\langle n \rangle_j$ has been calculated, the diffusion and subsequent immobilization of a second adatom on the array can be studied, subject to the constraint that once the first adatom has been immobilized at a particular site, that site is inaccessible to a second diffusing adatom (i.e., the site occupied by the first adatom is "blocked"). The $\langle n \rangle_j$ $(j = 2)$ corresponding to this second event in a j-stage process can be calculated, as can the $\langle n \rangle_j$ for each subsequent stage, all subject to the constraint that as the island morphology begins to unfold, sites occupied previously at that stage of pattern formation are inaccessible to subsequent diffusing adatoms.

Two features of the above description are worth emphasizing before proceeding. First, as sites become "blocked," the diffusion space accessible to a diffusing adatom is modified: not only do subsequent adatoms have available a (somewhat) smaller space within which to diffuse but, more importantly, they also have to negotiate "obstacles" (i.e., the blocked sites). Here, we consider first diffusion restricted to a basal plane only, and then diffusion in an expanded space, basal plane, and overlayer; this to account for the possible detachment from and redeposition on the basal surface. Blocked sites on the template are sterically excluded to an adatom diffusing in the overlayer (as well as on the basal plane).

Second, in a j-step deposition/docking process, depending on the symmetry of the island assembly, there are *at least* $j!$ different ways in which a final morphology can be realized (i.e., there are many developing configurations of blocked sites that can lead (eventually) to the same final pattern). In the simplest case, if a "square" island has been assembled and the deposition of a new j-site boundary layer is initiated, there are $j!$ different ways in which the sequential addition of adatoms can occur leading to the same final configuration. More important, in forming, say, a platelet of M sites, there are reaction pathways involving a manifold of

possible, geometrically distinct "intermediates" that can result in the same final morphology. Our calculations show that after the first few structural generations, the former combinatorial effect results in only small quantitative changes in the signatures $\langle n \rangle_j$, but evolution via *different* pathways to a given final configuration can lead to significant differences in the $\langle n \rangle_j$ values.

II. DESCRIPTION OF THE MODEL

We have developed a lattice statistical model and followed the evolution of specific patterns by obtaining numerically exact solutions to a stochastic master equation formulated for this problem. This strategy provides an approach complementary to the kinetic Monte Carlo or molecular dynamics simulations that have been reported [3]. The approach elaborated here has the advantage that once the problem is formulated, no further approximations, analytic or computational, are introduced in characterizing the dynamics [4].

The lattice model developed is $d > 2$ dimensional and may be thought of as stacked "checkerboards," two in number (see, however, later discussion). The basal layer is represented by a lattice of valence $\nu = 4$ or 6. The overlayer is characterized by the same planar symmetry as the basal lattice, but with an additional "bond" linking the overlayer to corresponding sites on the basal plane. For example, if the basal plane is characterized by square-planar symmetry, introducing an overlayer of square-planar symmetry generates a lattice each site of which is of valence $\nu = 5$; stacking triangular lattices generates an extended lattice, each site of which is of valence $\nu = 7$.

We imagine that an atom or molecule can initiate its motion at any site on the basal plane or at any overlayer site, subsequently undergoing random displacements until localized (or "trapped") at a lattice site on the basal plane. If the particle happens to be situated on a basal plane having square-planar symmetry, six motional degrees of freedom are assigned, namely, four in the plane, one vertical displacement to the overlayer (to account for detachment of adatoms from the substrate), and one "reset" (or "stall") at the same lattice site (to account, approximately, for surface forces). If the particle happens to be in the overlayer, we assign four motional degrees of freedom to account for lateral diffusion in the overlayer, allow a possible excursion to the basal plane (to characterize the adsorption or readsorption of the adatom), and impose a "reset" condition if the particle attempts to exit the system ("zero flux" boundary conditions). For both the basal plane and the overlayer, periodic boundary conditions are imposed when the adatom confronts the lateral boundaries of the system.

As noted, an adatom diffusing in the overlayer and attempting to exit the system is confined by a passive ("reset") boundary condition. Physically, this confinement may be a consequence of having an impenetrable barrier or as a

consequence of the region above the overlayer being "concentrated" with respect to similar diffusing species (thus simulating an unfavorable density gradient). Apart from the "injected" particle, the overlayer itself is free of other particles, so this region may be thought of as the lattice equivalent of a classical "depletion zone."

To summarize, each lattice site of a layered system having square-planar symmetry has an overall valence of $\nu = 6$, and each site on layered system having triangular symmetry has an overall valence of $\nu = 8$. In this formulation, sites on a $d = 2$ dimensional basal template of triangular symmetry have the same valence as sites on stacked lattices of square-planar symmetry, namely, $\nu = 6$, a circumstance that will permit later comparison of pattern formation on strictly $d = 2$ dimensional versus quasi-two-dimensional systems.

The formation of patterns is studied as follows. We specify a docking site at some location on the basal plane, and by solving the stochastic master equation, we determine the eigenvalue spectrum characterizing the evolution of the system until the diffusing particle is trapped (here) irreversibly at the basal site. From the theory of finite Markov processes [5], we know that the smallest eigenvalue dominates the long-time behavior of the system. We also know that in the limit of large lattices, the smallest eigenvalue is related to the reciprocal of the first moment of the underlying probability distribution function, the mean walklength $\langle n \rangle$, a relationship that becomes exact in the infinite system limit.

The mean walklength will be used here as a surrogate for the rate constant characterizing the formation of a specific pattern for two reasons. First, the lattices are large. The basal plane of both square-planar and triangular lattices was set at $N = 272$ sites, so each composite system has 544 sites. Explicit calculation shows that the error introduced in using the reciprocal of the mean walklength to portray the smallest eigenvalue is less than 1%. Second, the mean number of displacements of a diffusing particle before localization provides a more intuitive understanding of the event than reporting the values of rate constants (unless, of course, one is tying the calculation to a specific experimental study for which rate data are available). Note that specifying both lattice systems to have the same number N of total lattice sites allows a direct (and unambiguous) comparison of the consequences of adopting different lattice symmetries.

After the first "docking" event has occurred, the immobilization of a second adatom can be studied by following the fate of a new "injected" particle. Implementing this protocol M times, the development of arbitrary patterns on the host template can be explored. We shall consider here patterns having the morphology of thin dendrites, thick dendrites, and faceted platelets.

Finally, we comment briefly on the stochastic calculation itself. There are two checks on the accuracy of the calculations, which we illustrate for the case of a single trap. First, the imposition of periodic boundary conditions on the lateral faces of the lattice demands that the value of $\langle n \rangle$ calculated should be *invariant*

with respect to the position of the trap on the basal plane. Second, for lattices of uniform valence ν, Montroll [6] proved that in any Euclidean dimension, the mean walklength from sites nearest neighbor to the trap is given by the *integer* $\langle n \rangle$ $(\nu) = N - 1$, a result that is satisfied numerically if (and only if) the values of $\langle n(i) \rangle$ are numerically *exact*. These two criteria for the accuracy of our Markovian calculations have been confirmed for all cases studied here.

III. ISLAND GROWTH IN $d = 2$

First, we consider the stochastic consequences of adopting different scenarios in the self-assembly of a specific nucleation pattern on a $d = 2$ dimensional template. In an earlier study [7, 8], arrays having square-planar or triangular symmetry were considered, with the total number N of sites defining the host lattice fixed at $N = 48$. The consequences of following different pathways leading to the formation of an $M = 16$ platelet on the host lattice were studied quantitatively; these pathways defined possible modes of island growth via a "row filling," a "dendrite," or a "compact" mechanism.

The present work enlarges the program elaborated in Refs 7, 8 in three specific ways. First, more robust patterns were considered, namely, dendrites having both thin and thick branches and larger platelets. In each case, the final assembled unit has $M = 37$ sites; hence, much larger lattices were necessary to support an expanded diffusion space (see earlier specification). Second, only confined systems were studied only in Ref. 8; that is, self-assembly was studied on domains of finite extent. Here, we impose periodic boundary conditions specifically to avoid considering "boundary effects." Third, the study [8] was restricted to patterns assembled on a basal plane only; here, the diffusion space is augmented by the presence of an overlayer (see following section).

The patterns generated via sequential atom-by-atom deposition are illustrated and specified in Figs. 1 and 2. For each lattice geometry, three pathways are studied, two of which lead to the same, final pattern of $M = 37$ sites (Fig. 1a and b), and the third that has the same number ($M = 37$) of total sites, but a different final morphology (Fig. 1c). These choices were motivated by the interest in studying the relative efficiency of two pathways that lead to the same final localized pattern (a platelet) versus a third pathway that leads to a dendrite having a fractal-like distribution of deposited adatoms.

We focus first on the relative efficiency of two pathways, both of which lead to an eventual, common pattern, a platelet. The $\langle n(i) \rangle$ values for platelet formation generated via a "boundary layer" growth mechanism versus a "thick dendrite" growth mechanism on templates of triangular versus square-planar symmetry can be compared. To make the comparison more transparent, we normalize the results obtained at each stage of growth by the $\langle n \rangle$ value for a single trap on a $d = 2$

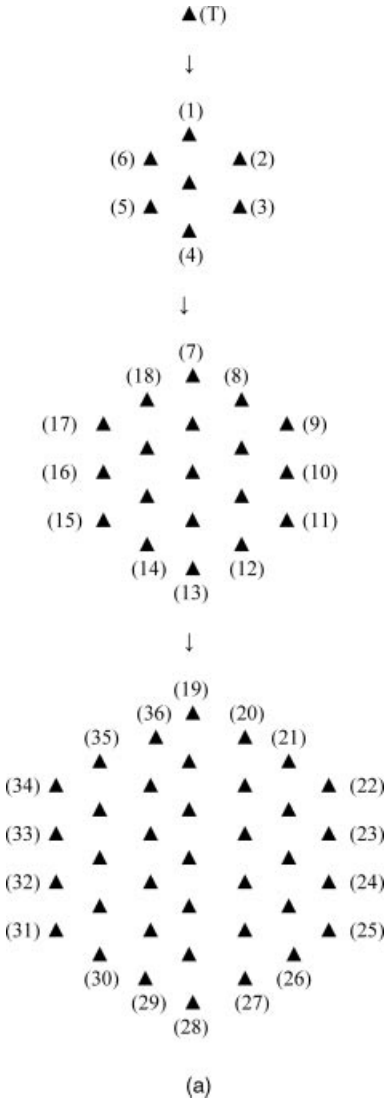
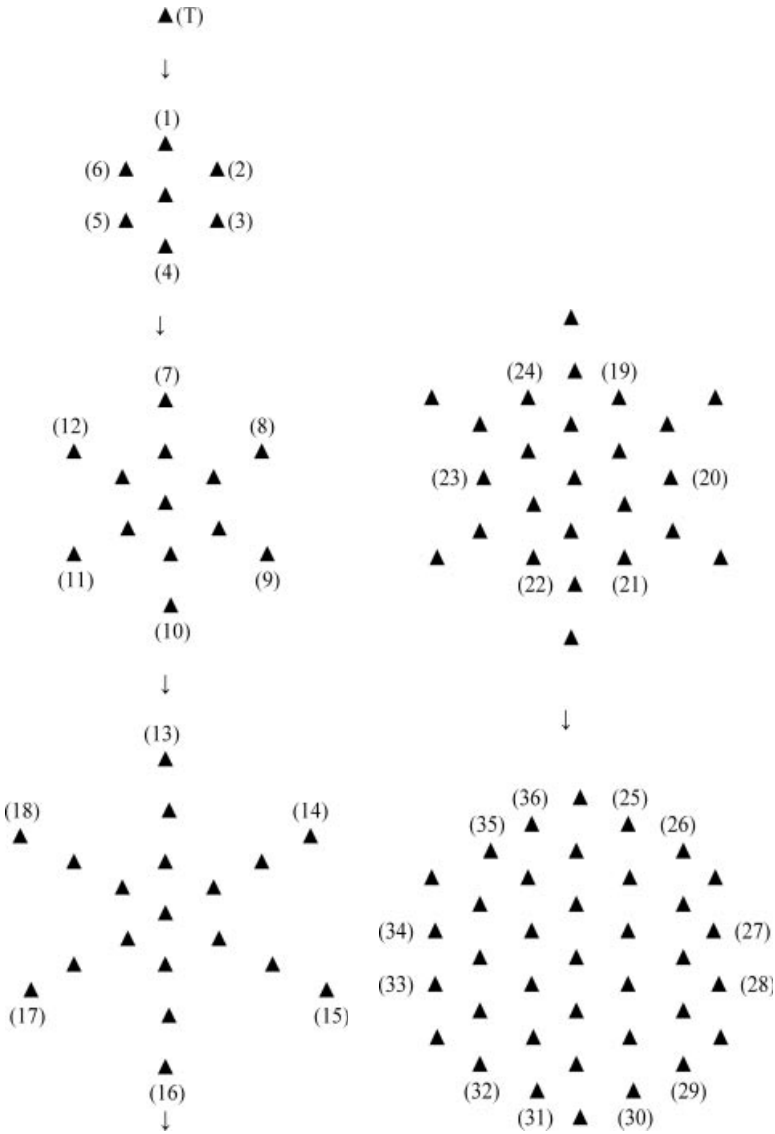


Figure 1. (**a**–**c**) Three patterns generated via a sequential atom-by-atom deposition on a triangular lattice. Here, we denote (a) platelet, (b) thick dendrite, and (c) thin dendrite.

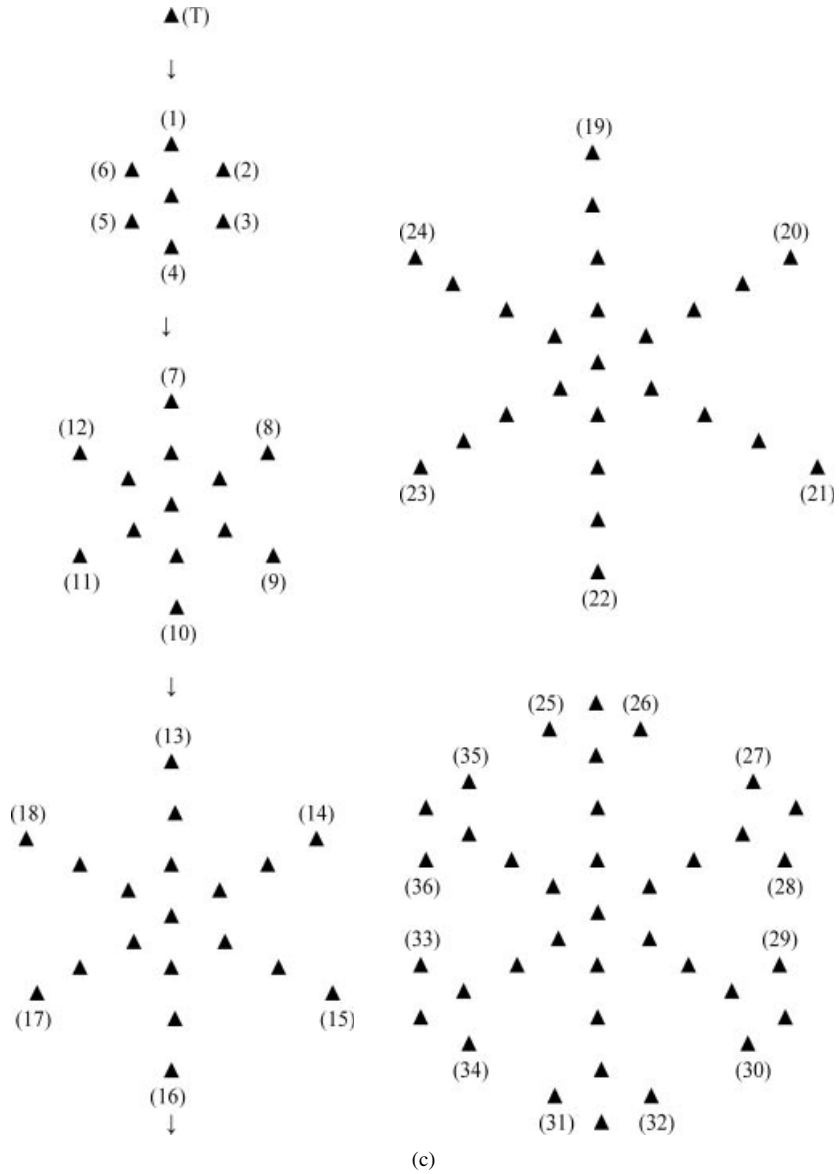
dimensional lattice of the same size and symmetry and subject to the same boundary conditions. We denote the normalized metric, $\langle n \rangle / \langle n \rangle_{\text{single trap}}$, by η.

Figure 3a compares the growth histories for scenarios shown in Fig. 1a and b for lattices of triangular symmetry, and Fig. 3b compares the growth history diagrammed in Fig. 2a and b for lattices of square-planar symmetry. Then, in Fig. 4a, the relative efficiency of "thick" dendrite growth versus "thin" dendrite growth



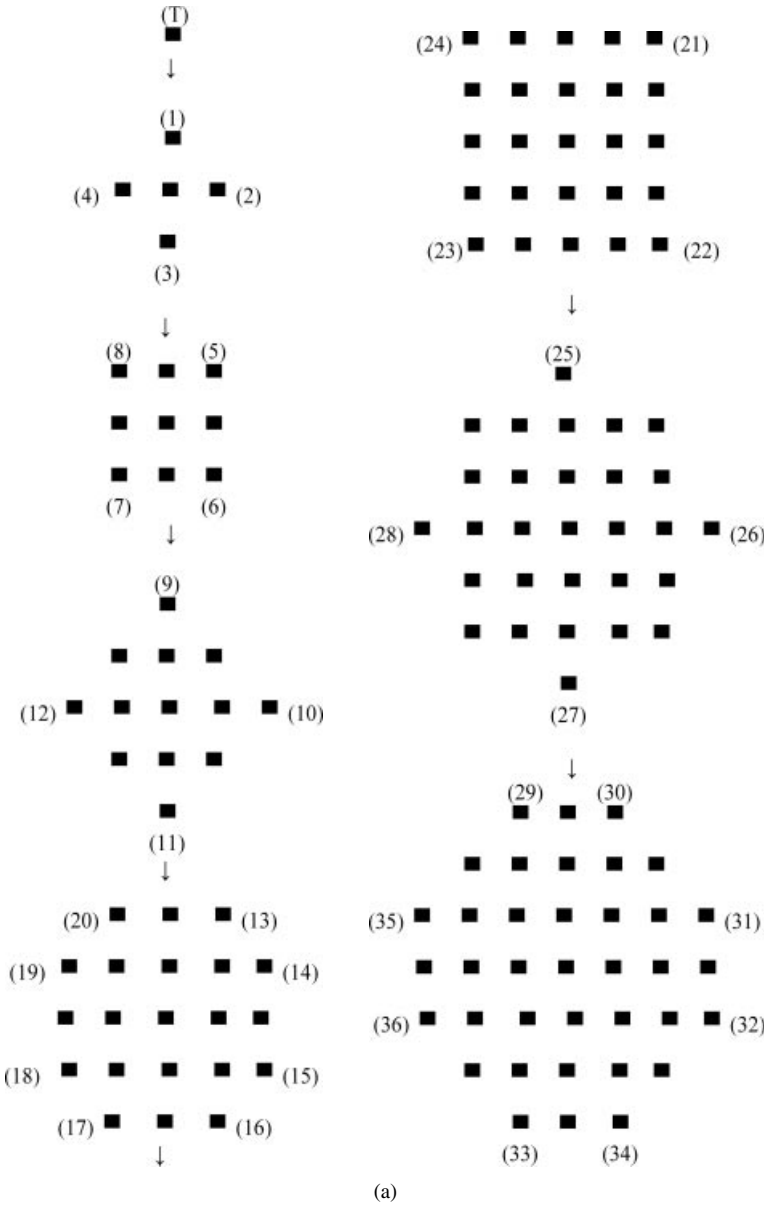
(b)

Figure 1. (*Continued*)



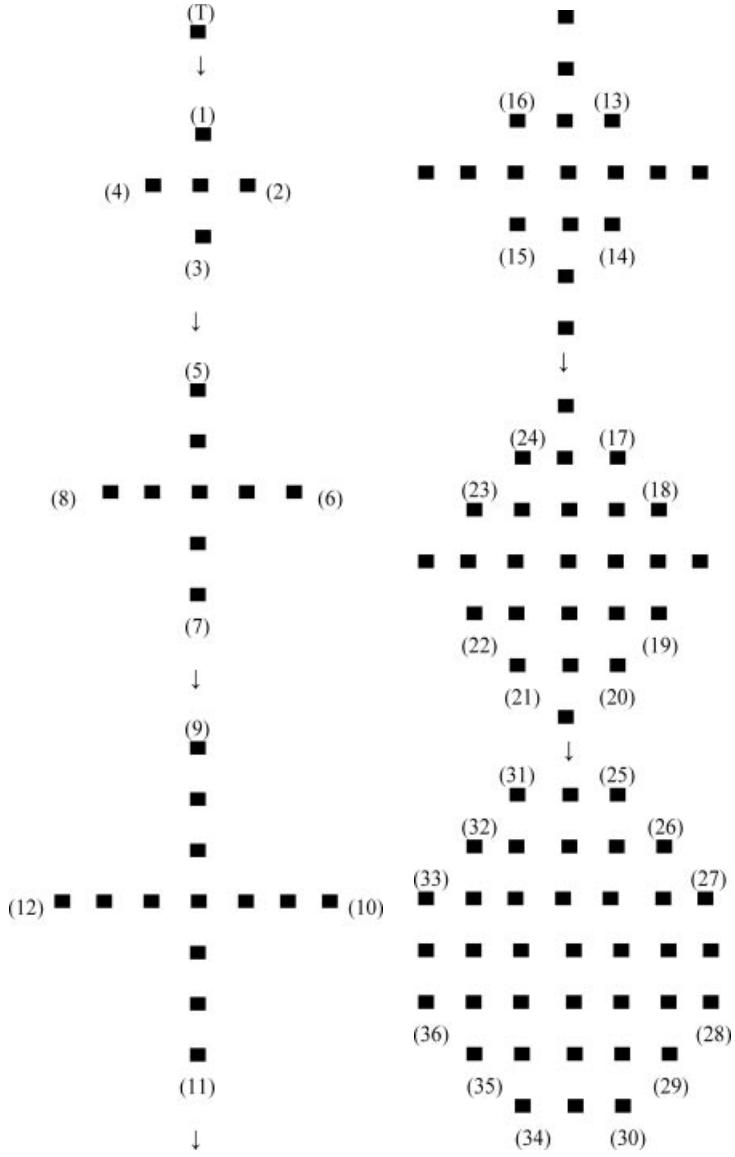
(c)

Figure 1. (*Continued*)



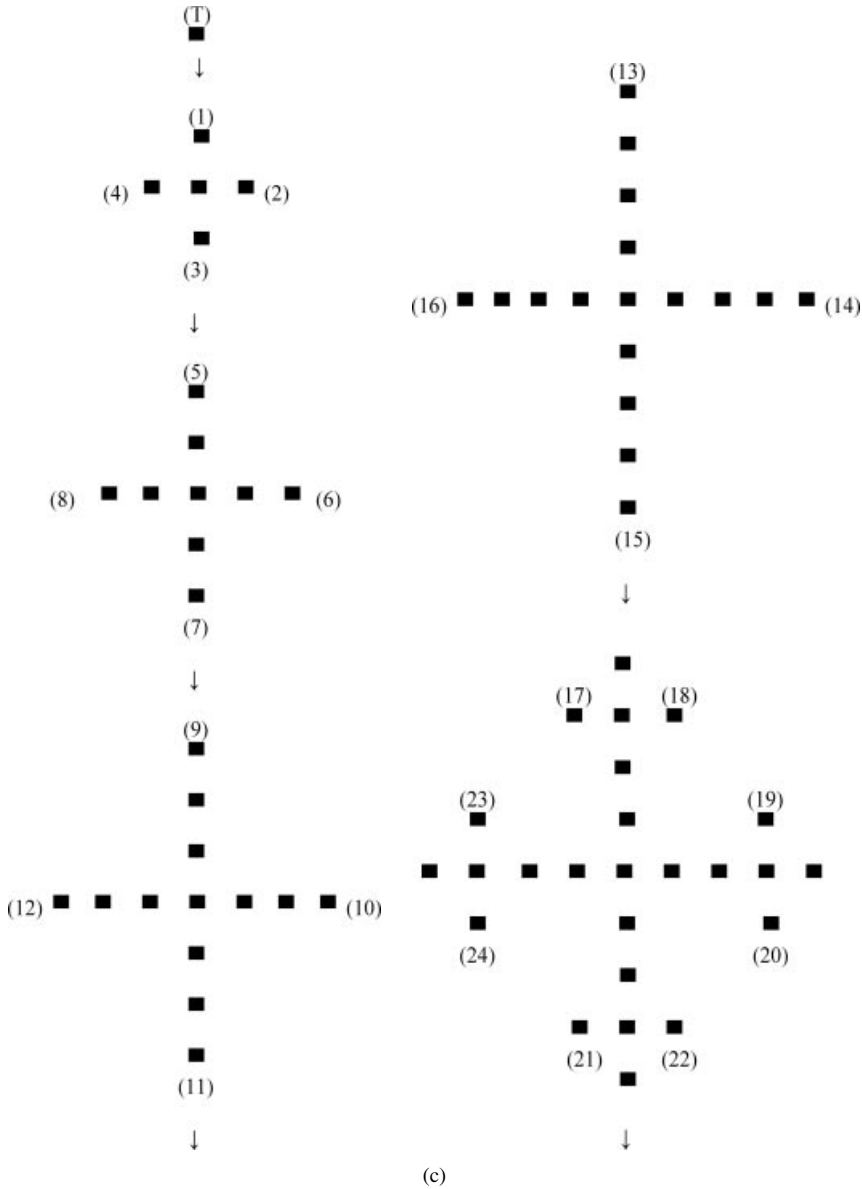
(a)

Figure 2. Three patterns generated via a sequential atom-by-atom deposition on a square-planar lattice. Here, (**a**–**c**) are as specified in Fig. 1.



(b)

Figure 2. (*Continued*)



(c)

Figure 2. (*Continued*)

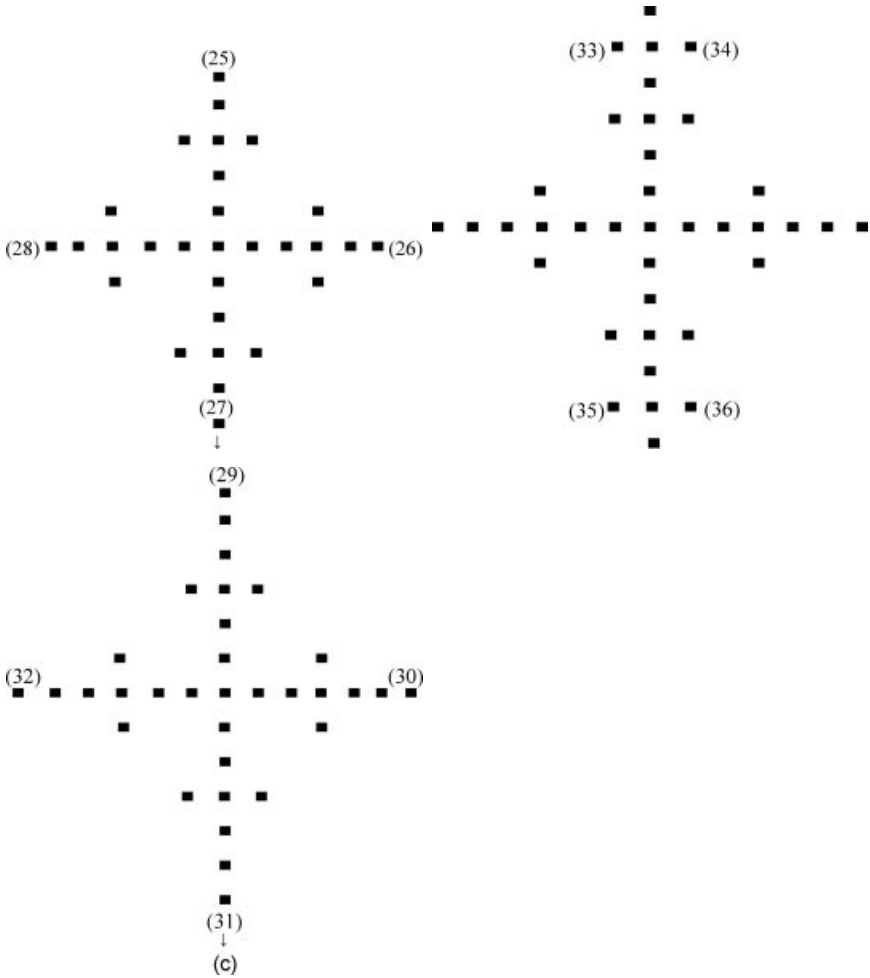


Figure 2. (*Continued*)

on lattices of triangular symmetry is compared; a similar comparison is shown in Fig. 4b for lattices of square-planar symmetry. Again, for emphasis, we note that, in our model, both the triangular and square-planar lattices have the same number N of lattice sites ($N = 272$) and the same number M of sites defining the final configuration generated ($M = 37$). The results presented in Figs. 3 and 4 will be discussed in Section V.

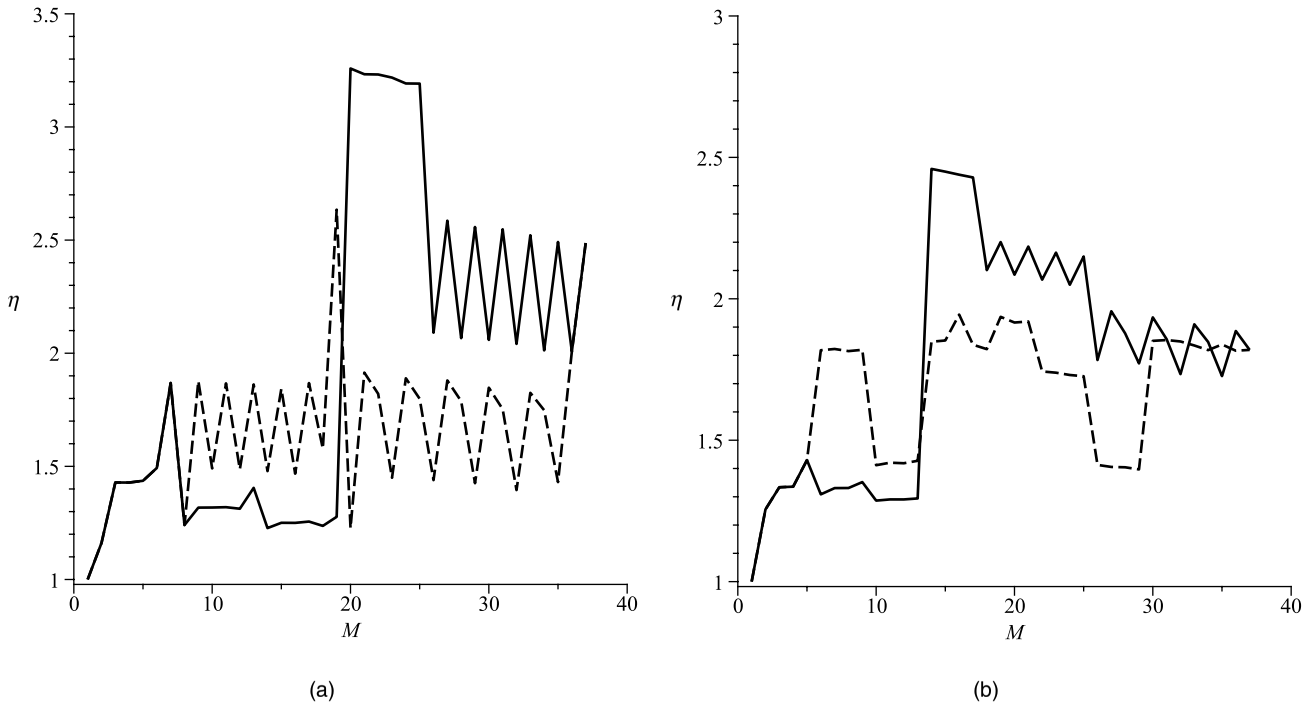


Figure 3. (**a**) Comparison of growth history for scenarios shown in Fig. 1a and b on a basal template of triangular symmetry. Vertical axis is η, the walklength $\langle n \rangle$ is normalized with respect to the $\langle n \rangle$ value for a single trap, and the horizontal axis is the number M of (sequentially added) adatoms. Platelet evolution is denoted by dashed line and thick dendrite evolution by a solid line. (**b**) Comparison of growth history for scenarios shown in Fig. 2a and b on a basal template of square-planar symmetry. Same axes and line conventions as Fig. 3a.

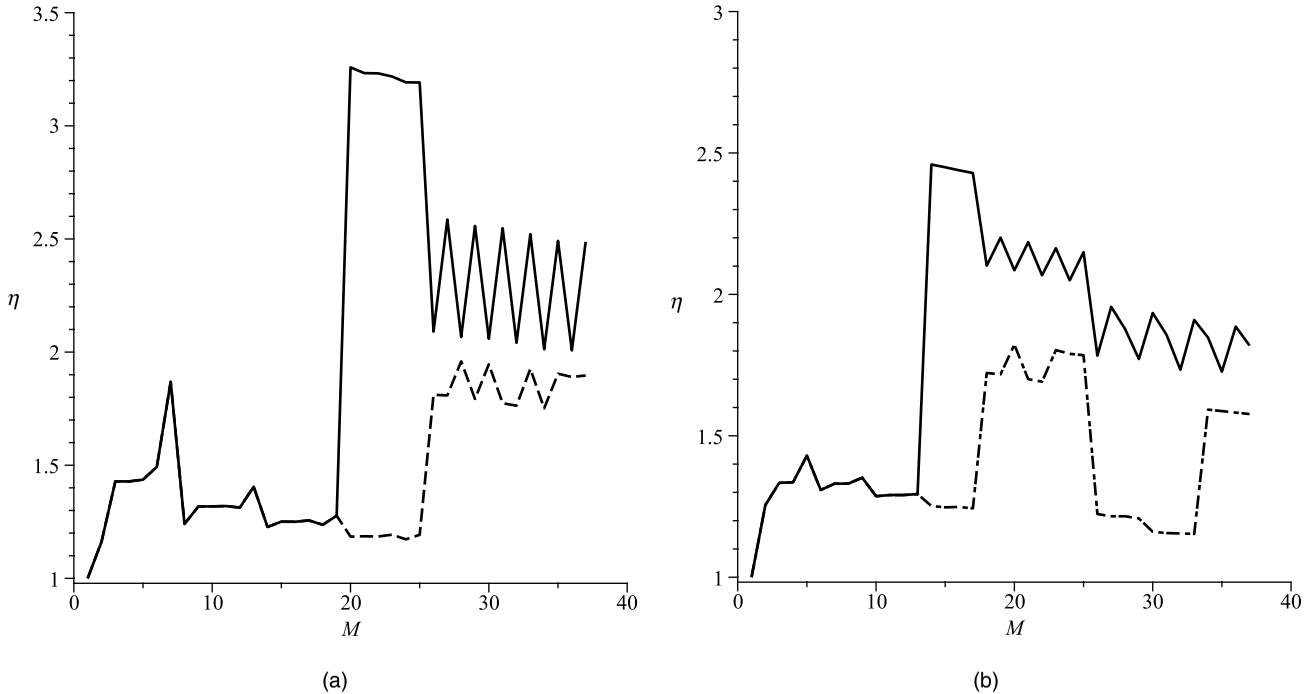


Figure 4. (**a**) Comparison of growth history for "thin" versus "thick" dendrite formation on a basal template of triangular symmetry. Same axes convention as in Fig. 3a. Thin dendrite evolution is denoted by a "dot-dash" line and thick dendrite evolution by a solid line. (**b**) Comparison of growth history for "thin" versus "thick" dendrite formation on a basal template of square-planar symmetry. Same axes and line conventions as in Fig. 4a.

IV. ISLAND GROWTH IN $d = 3$

To study the consequences of expanding the diffusion space, a program of calculation companion to that reported in Section III can be carried out by considering a composite system, basal plane plus overlayer. Figure 5a and b may be placed in tandem with Fig. 3a and b, and (a) and (b) parts of Fig. 6 are the counterparts of Fig. 4a and b. In Figs. 5 and 6, the value of $\langle n \rangle$ used to normalize the walklength data is the mean walklength for a walker on the composite lattice (of given symmetry) with a single trap on the basal plane. These results and the conclusions drawn will be discussed in the following section.

We conclude this section with a discussion of the consequences of expanding further the diffusion space of the system, specifically by adding additional overlayers. One anticipates that the quantitative results will change (walklengths will become longer), but the more interesting question is whether, and to what extent, the qualitative behavior remains unchanged as the diffusion space is expanded.

Calculations were performed for the three scenarios shown in Fig. 1 and for a template positioned on layer 3 of a five-layer expanded lattice. Each layer of this expanded lattice has triangular symmetry, and each site is of valence $\nu = 8$, with the overall number N of sites equal to $N = 5 \times 272 = 1360$ sites. As expected, the mean walklength at each stage of self-assembly of a given pattern is numerically much larger, a factor of 5 or greater.

To compare the *qualitative* results obtained for diffusion-reaction processes on a composite lattice having a single overlayer ($N = 524$ sites) versus those generated for a five-layer composite system ($N = 1360$ sites), we normalize (as before) the calculated $\langle n(i) \rangle$ values in each case by the $\langle n \rangle$ value for a single trap positioned on the given host template. As shown in Fig. 7, for each of the three scenarios diagrammed in Fig. 1, and for both composite lattices, the qualitative behavior is remarkably the same. We conclude that the qualitative results obtained by adding multiple overlayers are already captured by those obtained considering a single overlayer.

V. DISCUSSION

We now review the principal results obtained in this work. Recall that for large lattices, the smallest eigenvalue obtained in solving the underlying stochastic master equation is safely approximated by the inverse of the mean walklength $\langle n \rangle$. In effect, the larger the mean walklength, the slower the chemical rate of deposition.

Focusing first on self-assembly in strictly $d = 2$ systems, and using the mean walklength as a diagnostic tool, several conclusions follow from examination of the results shown in Figs. 3a and 4a, which compare final platelet formation as generated via two different pathways, a "boundary-deposition" pathway versus a

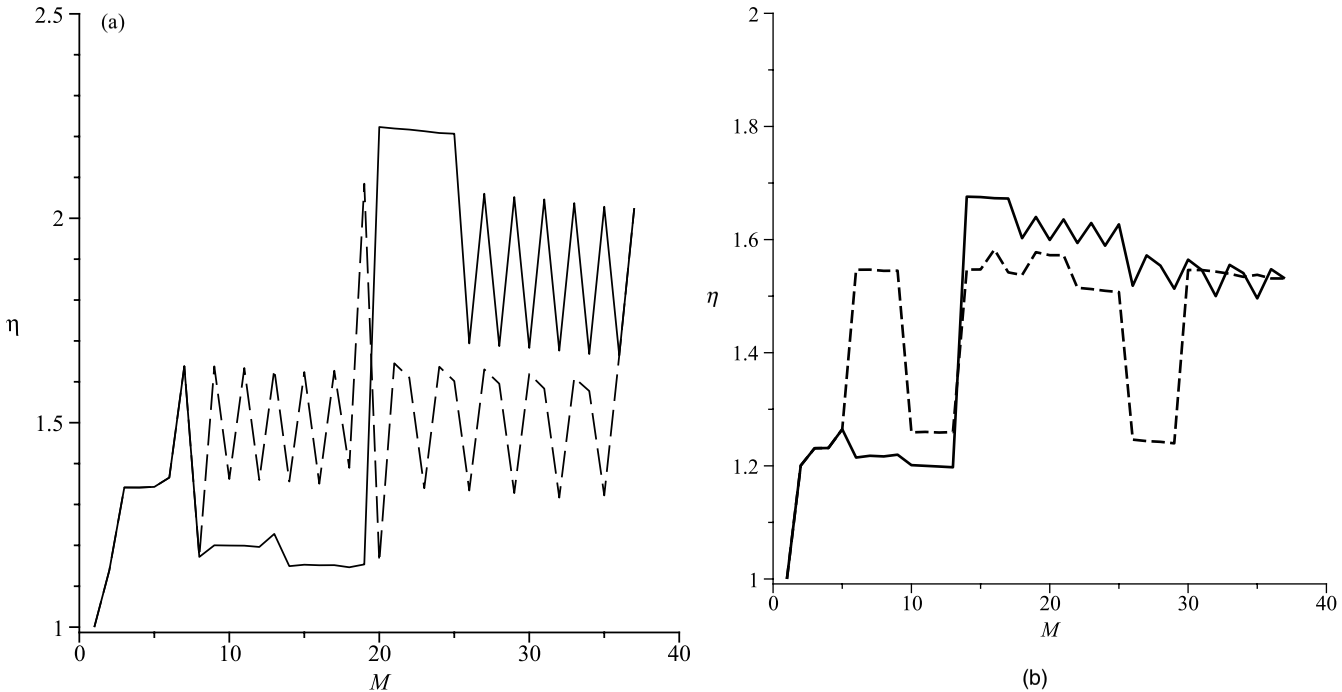


Figure 5. (**a**) Comparison of growth history for scenarios shown in Fig. 1a and b on a basal plane (solid line) versus basal plane plus overlayer (dashed line) of triangular symmetry. Same axes conventions as in Fig. 3a. (**b**) Comparison of growth history for scenarios shown in Fig. 2a and b on a basal plane (solid line) versus basal plane plus overlayer (dashed line) of square-planar symmetry. Same axes conventions as in Fig. 3a.

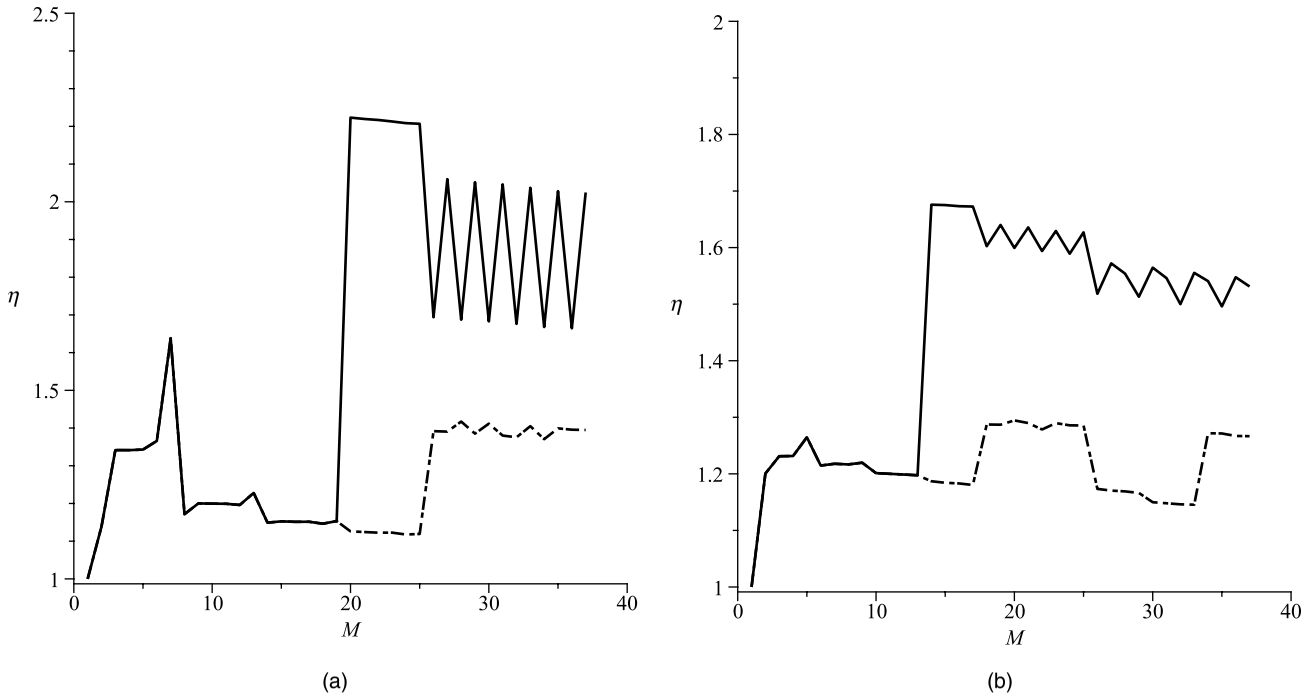


Figure 6. (**a**) Comparison of growth history for "thin" versus "thick" dendrite formation on a basal plane (solid line) versus basal plane plus overlayer (dashed line) of triangular symmetry. Same axes conventions as in Fig. 3a. (**b**) Comparison of growth history for "thin" versus "thick" dendrite formation on a basal plane (solid line) versus basal plane plus overlayer (dashed line) of square-planar symmetry. Same axes conventions as in Fig. 3a.

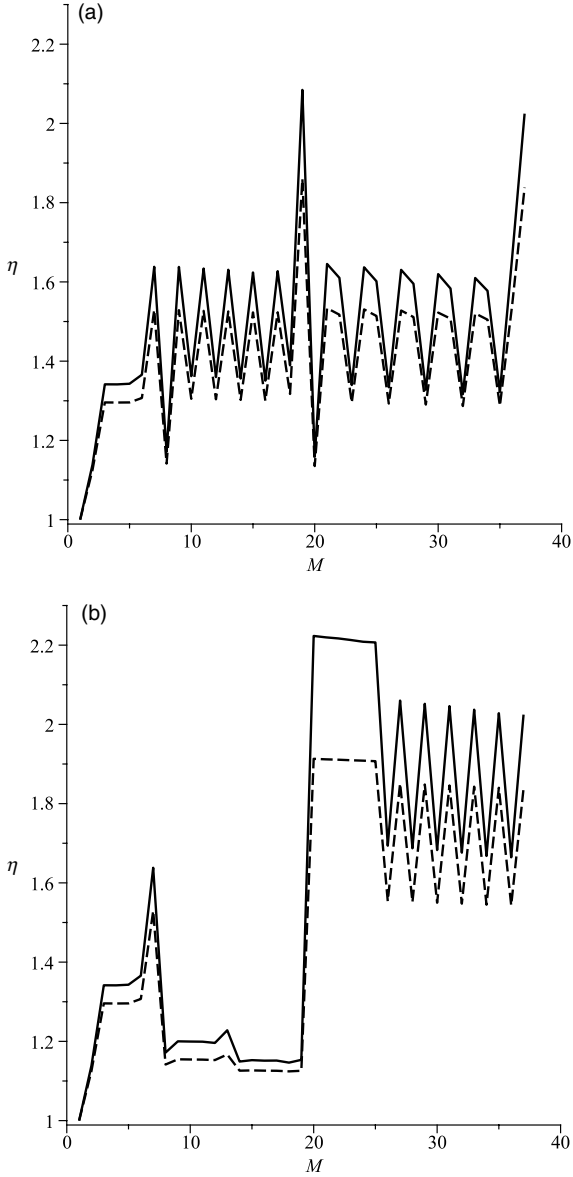


Figure 7. Comparison of growth history for the three growth scenarios diagrammed in Fig. 1a–c for a composite system having single overlayer (solid line) versus a system with five overlayers (dashed line) with deposition in the latter on the central plane. Each set of results is normalized with respect to the $\langle n \rangle$ value for a single trap; same axes conventions as in Fig. 3a. (**a**) Platelet; (**b**) thick dendrite; (**c**) thin dendrite.

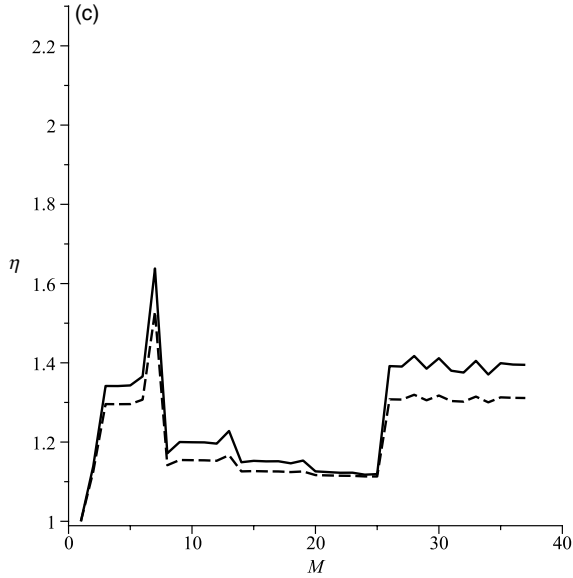


Figure 7. (*Continued*)

"thick dendrite" pathway, on lattices of triangular and square-planar symmetries, respectively.

The most obvious point to be made about the results shown in Figs. 3a and 4a is that evolution to the final platelet structure (shown in Fig. 1a and b and Fig. 2a and b) is far from uniform; the self-assembly is neither smooth nor continuous. Translating this observation into the language of chemical kinetics, the rate constants characterizing particular stages in the evolution change, *discontinuously*, as adatoms are deposited sequentially, a reflection and consequence of the site-specific local geometry. It is noteworthy that there are sequences where boundary deposition is more efficient (higher rate constants) than thick dendrite formation, and vice versa. Empirically, one finds that "crests" in the "sawtooth" patterns generated correspond to deposition at "notches" in a developing growth pattern, whereas the "troughs" correspond to deposition at boundary sites flanking the "notch." While an interesting observation, it is clear that to compare the overall efficiency of competing pathways in a more systematic way, one needs to go beyond a simple visual scan of the displayed profiles; a diagnostic "index" is needed to quantify differences. We shall return to this point later.

The profiles shown in Figs. 3b and 4b are much easier to interpret. In these plots, we compare thick versus thin dendrite growth, eventually leading to a final ($M = 37$) assembly having two distinct morphologies (Figs. 1b and c and 2b and c).

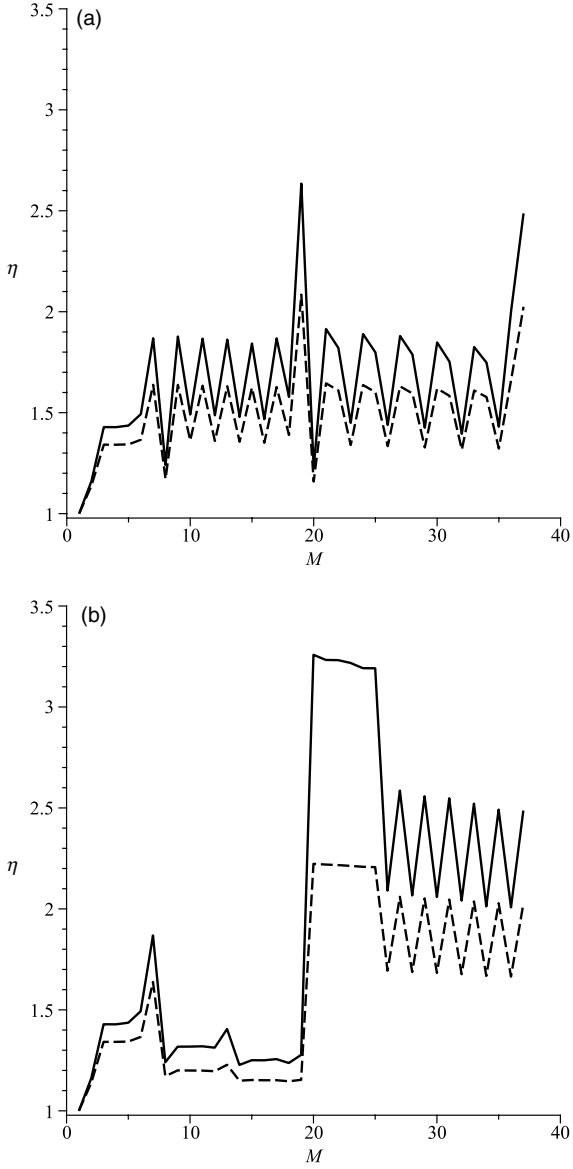


Figure 8. (**a**) Consequences of adsorption/desorption in platelet formation (Fig. 1a). Growth on a basal plane (solid line) versus basal plane plus overlayer (dashed line), both of triangular symmetry. (**b**) Consequences of adsorption/desorption in thick dendrite formation (Fig. 1b). Growth on a basal template versus basal plane plus overlayer, both of triangular symmetry. Same convention as in Fig. 8a. (**c**) Consequences of adsorption/desorption in thin dendrite formation (Fig. 1c). Growth on a basal template versus basal plane plus overlayer, both of triangular symmetry. Same convention as in Fig. 8a.

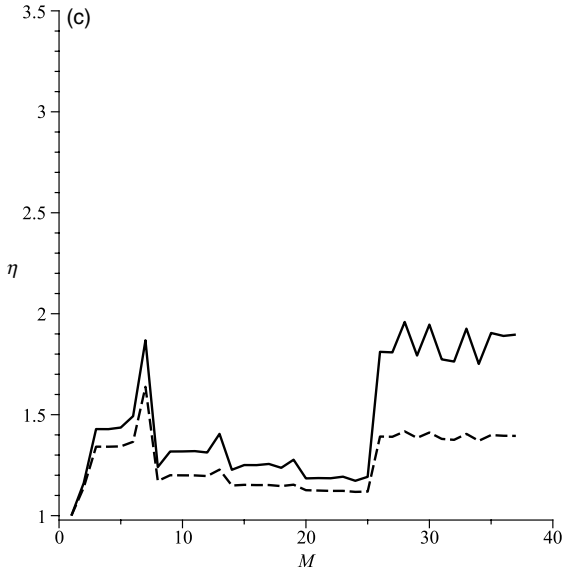


Figure 8. (*Continued*)

Formation of thin dendrites is characterized by faster growth. In qualitative language, the more distributed the set of deposited adatoms, the more efficient is self-assembly to a final pattern. A similar conclusion was reached and documented extensively in an earlier study of diffusion-reaction processes on $d = 2$ templates [9, 10]. Turning next to $d > 2$ dimensional composite lattices (basal plane and one overlayer), we find that the results shown in Figs. 5a and 6a for systems of triangular and square-planar symmetry, respectively, are qualitatively the same as those obtained when diffusion-reaction processes on a basal plane only are considered. This similarity can be seen most convincingly by examining Fig. 8 for lattices of triangular symmetry and Fig. 9 for lattices of square-planar symmetry. In these plots, (normalized) walklength data are compared directly for $d = 2$ versus $d > 2$ lattices for each pathway diagrammed in Figs. 1 and 2. Also of interest are the results presented earlier, Fig. 7, for multilayered lattices of triangular symmetry. When the latter are compared with the results shown in Fig. 8, we find that the qualitative similarities/differences among different pathways of self-assembly in multilayered systems are already captured by the diffusion-reaction behavior displayed on a basal template (only).

All the above results have been obtained for strictly entropic processes, namely, no attractive forces are at play in influencing the pattern formation. In his benchmark experimental and analytical studies on the physics of snow crystals, Libbrecht

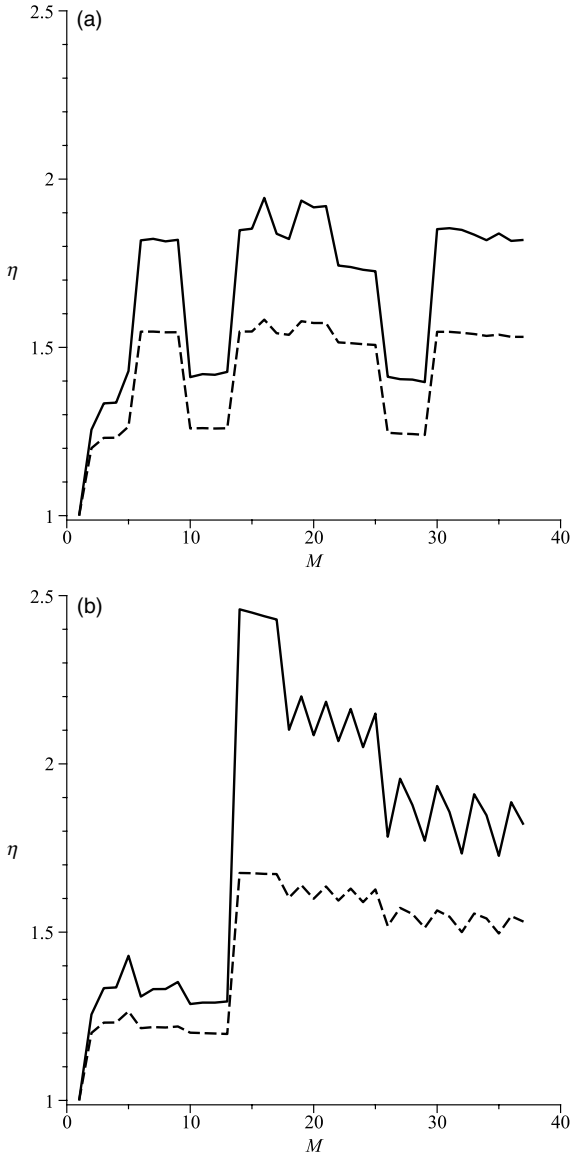


Figure 9. (**a**) Consequences of adsorption/desorption in platelet formation (Fig. 2a). Growth on a basal template versus basal plane plus overlayer, both of square-planar symmetry. Same convention as in Fig. 8a. (**b**) Consequences of adsorption/desorption in think dendrite formation (Fig. 2b). Growth on a basal template versus basal plane plus overlayer, both of square-planar symmetry. Same convention as in Fig. 8a. (**c**) Consequences of adsorption/desorption in thin dendrite formation (Fig. 2c). Growth on a basal template versus basal plane plus overlayer, both of square-planar symmetry. Same convention as in Fig. 8a.

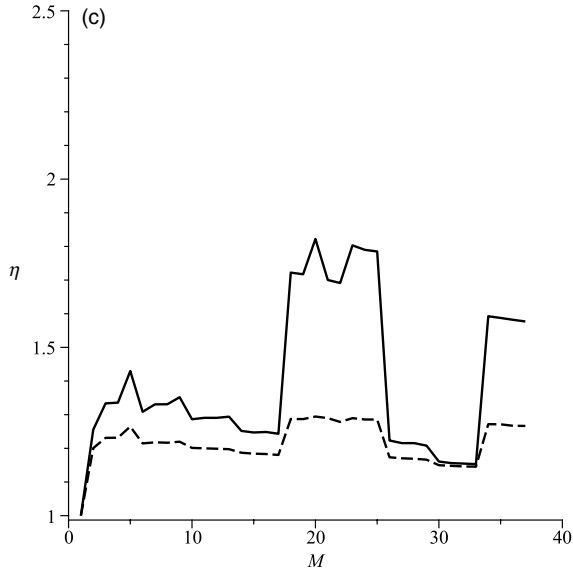


Figure 9. (*Continued*)

[11] has emphasized the sensitivity of the deposition process to the "condensation coefficient" α at particular sites. Libbrecht notes that α embodies the surface physics that governs how water molecules are incorporated into a host lattice. He notes [11] that "α will depend on T, σ_{surf}, the surface structure and geometry, surface chemistry, etc.," where T is temperature and σ_{surf} is the supersaturation just above the growing surface. "If molecules striking the surface are instantly incorporated into it, then $\alpha = 1$; otherwise we must have $\alpha \leq 1$." More recently, the sensitivity of the "docking" efficiency to a range of values of an "attachment coefficient" ($0 \leq \alpha \leq 1$) has been documented in a model of the early stages of self-assembly in nanophase materials [12, 13].

The sensitivity of the results obtained here to the coefficient α can be illustrated by focusing on site (19) in Fig. 1b versus site (19) in Fig. 1c; both sites are on a basal plane of triangular symmetry. In a purely entropic process, attachment at site (19) in Fig. 1c is favored over attachment at site (19) in Fig. 1b. The results shown in Fig. 10 show the consequences of increasing α from its "random walk" value, $\alpha = 1/6$, to $\alpha = 1$. There, the $\langle n \rangle$ value for localization at site (19) in Fig. 1c (horizontal line) and the $\langle n \rangle$ values for attachment at site (19) in Fig. 1b for a range of α values can be compared. Even setting $\alpha = 1$, attachment at site (19) in the "thin" dendrite, Fig. 1c, is still favored, suggesting that steric factors play the more decisive role in island growth.

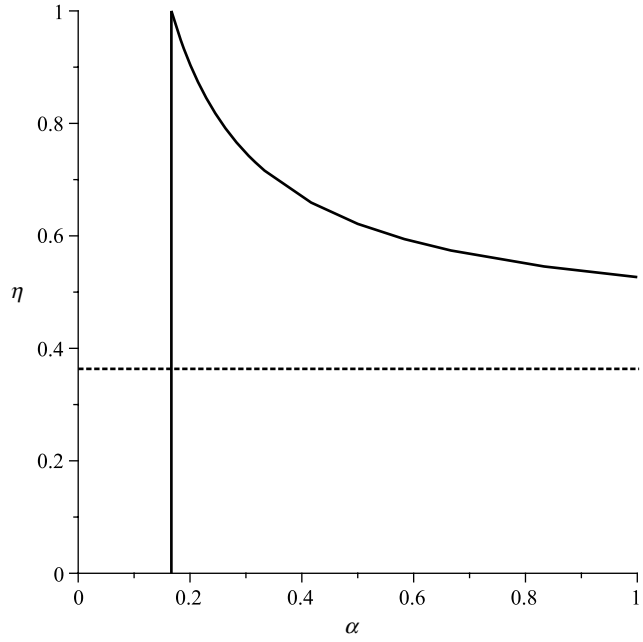


Figure 10. Dependence of localization of a diffusing adatom at site (19) in Fig. 1b on the condensation coefficient α . The vertical axis is specified as in Fig. 3a. The horizontal axis specifies α . The horizontal line corresponds to the value of $\langle n \rangle$ for site (19) in Fig. 1c. All $\langle n \rangle$ values are normalized with respect to the $\langle n \rangle$ value for the setting $\alpha = 1/6$ for the case of Fig. 1b.

VI. CONCLUSIONS

Remarkable patterns ranging from fractal-like dendrites to platelets can be formed by minerals deposited on surfaces. Among the many examples that can be cited, we draw attention to patterns formed by the hollandite group of manganese oxides [14–16] in, for example, desert varnishes, and the recently discovered "snowflake" patterns formed in zinc oxide synthesis [17]. Chopard et al. [18] have emphasized that "mineral dendrites are distinct from the 'dendritic' crystal morphologies obtained by solidification from an undercooled melt." The latter class of problems has been examined extensively in the literature [1], with notable contributions by, among others, Langer [19] and Viscek [20]. Chopard et al. also argue that diffusion-limited aggregation (DLA) [21] "cannot explain all features of mineral dendrites," pointing out that the chemistry is more complex, and "observed dendrites have varying fractal dimensions, whereas DLA results in a unique value $D \sim 1.70$." A

related point is made by Zhang and Lagally [2d] in their discussion of thin film growth. They note that "in the DLA model, the average branch thickness b of the fractal island is about one atom wide ($b \sim 1$), while the fractal islands observed with the STM all have wider branch thicknesses" (see also Refs 22–27).

Chopard, Herrmann, and Vicsek (CHV) developed a diffusion-reaction model [18] to study the structure and growth of mineral dendrites, with particular reference to the oxidation of Mn^{4+} ions diffusing out of cracks (e.g., in limestone) and combining with O^{2-} to form MnO_2 (the mineral pyrolusite). We note that Potter and Rossman [14] pointed out (nearly 30 years ago) that "although dendrites have long been considered to be pyrolusite, no example of pyrolusite mineralogy has been identified." Recognizing that the hollandite group of minerals (hollandite, $BaMn_8O_{16}$, cryptomelane, KMn_8O_{16}, coronadite, $PbMn_8O_{16}$) are chemically distinct from pyrolusite, MnO_2, the *simulations* presented in Ref. 18 are nonetheless still of considerable interest. There, a square-planar lattice was used to describe the diffusion space. Two species, A and B, diffuse randomly until encountering each other at a particular site. The resulting product, C, subsequently diffuses (randomly) until precipitating ($C \rightarrow D$) at a particular site and is thereafter localized.

The authors note that two conditions will lead C to being deposited, that is, transformed to D (designated the "black" particle in their study): "first, when at least k particles of type C simultaneously meet on the same site, saturation occurs, and the C particles precipitate to form D; second, when a C particle becomes nearest neighbor to a black site D, it aggregates to the cluster."

With respect to the first of these deposition scenarios, numerically exact results generated via a Markovian theory, complementary Monte Carlo simulations and a kinetic analysis [28, 29], allowed a quantitative assessment of the relative efficiency of *synchronous* versus *asynchronous* motion of two diffusing reactants. As the number of reactants increases, the probability of *simultaneous* encounter of these species decreases. In the CHV model, the authors set $k = 4$, which is a relatively rare event, even in concentrated systems. The second deposition scenario proposed in Ref. 18 (deposition of C upon encountering a stationary D) is much more probable and also closely linked to models of thin film growth [2]. For these reasons, we have focused attention in this study on the second growth mechanism identified in the CHV model.

Our study was designed to explore the geometrical (steric) constraints governing and delimiting the final deposition of D. In addition to examining reaction pathways leading to thin or thick dendrites, we also considered platelet formation, the intent being to compare and analyze the relative efficiency of different diffusion-reaction schemes leading to a specific final self-assembled unit. The consequences of having available different reaction pathways in the self-assembly of nanophase systems have also been explored recently using a Markovian model [12, 13], and it was

documented that the observed kinetics depends decisively both on the number of reaction channels leading to a final self-assembled unit and (as here) on the relative efficiency of each of the competing reaction channels.

Turning next to the work of Gao and coworkers [17], these authors augmented their experimental study of "snowflake" formation in ZnO synthesis by carrying out Monte Carlo simulations in the "nonequilibrium growth regime." Although details of their simulations were not provided (nor any reference to the comprehensive study of Gravner and Griffeath [30]), they state that "periodical conditions were used and particles were added one by one into the region." Qualitatively, this scenario is similar to the one implemented here, and we will examine whether the results obtained in their Monte Carlo study are supported by the exact numerical results reported in previous sections. In their simulations, Gao and coworkers [17] distinguish between attachment at a kink site versus at "the straight part of the island edge," with the former characterized by $\alpha = 1$ and the latter by a probability p, or $\alpha < 1$. Gao and coworkers believe that "the dendritic patterns are mainly modulated by the total coverage," giving the following intuitive reason: "When the coverage is low, there is plenty of void space between the islands. Most of the particles attaching to the islands reach the islands from the outside "shores" rather than in the 'bays'." Our results show that even for a *single* island, geometric constraints already dictate the relative rate of formation of self-assembled units having different morphological structures, even when the attachment coefficient assumes the maximal value $\alpha = 1$. The "sawtooth" character shown in Figs. 3–9 dramatizes the relative efficiency of attachment to (in their language) "main branches" versus "side branches."

Whereas the relevant symmetry for ZnO "snowflake" formation is triangular, the "building blocks" for the hollandite group of manganese oxides are "blocks," as illustrated in Ref. 31. Considering dendrite formation for these minerals, reference to the results for lattices of square-planar symmetry is more relevant. Differences in the growth rate of two dendrite patterns here, "thick" and "thin," support and exemplify the importance attached to differences in the fractal dimension of the patterns simulated in Ref. 18. Chopard et al. comment that although their "numerically obtained pictures are very similar to the photographs of mineral dendrites, the underlying square lattice is still recognizable in the simulated precipitates. Therefore, off-lattice simulations would be interesting." The results presented here, especially Fig. 9 for diffusion-reaction processes on $d = 2$ versus $d > 2$ composite lattices of square-planar symmetry, argue that the main qualitative features of dendrite formation are already exhibited on $d = 2$ templates (see also the following discussion).

We now return to a point made earlier, the desideratum of characterizing in a more quantitative way the differences noted in, for example, the different profiles shown in Figs. 3–6. Perhaps the simplest signature that can be identified to characterize the differences is to compute and compare the area under each of

TABLE I
Normalized Areal Values Corresponding to Cases Considered in This Study

Pattern	Symmetry	Platelet	Thick dendrite	Thin dendrite
Basal Layer	Triangular	1.12	1.32	1.00[a]
	Square planar	1.19	1.28	1.00[b]
One Overlayer	Triangular	1.17	1.25	1.00[c]
	Square planar	1.18	1.22	1.00[d]

[a] Area = 52.667.
[b] Area = 50.109.
[c] Area = 45.322.
[d] Area = 43.705.

the curves generated.[1] If, then, one normalizes the results obtained to the smallest area calculated, one has a simple measure of the relative efficiency of "thin" versus "thick dendrite growth, versus "platelet" growth. Data for the different scenarios considered in this chapter are shown in Table I.

Using the calculated ratios to assess the overall (relative) efficiency of the three growth scenarios, one finds that "thin" dendrite formation is the most favorable pathway, followed by the "platelet" formation, and, finally "thick" dendrite formation. This conclusion is valid for both lattice geometries considered (basal layer (only) versus basal layer plus overlayer) and for both lattice symmetries (triangular and square planar). We also find that the ratio differences tend to be somewhat smaller for the overlayer calculation than for the basal layer only. In fact, if one carries out the areal calculation for the five-layer composite lattice described earlier, the differences are less than 5% among all three growth scenarios. The latter observation is a consequence of the fact that for very large lattices, the diffusing adatom spends most of its time diffusing in the overlayer; the structural differences among the three morphologies are a secondary factor. Conversely, geometrical differences among the patterns generated at each stage of self-assembly assume their greatest importance when diffusion-reaction processes are restricted to the basal plane only.

Given the design of our model, we noted earlier that by choosing a $d = 2$ basal plane of triangular symmetry and a $d > 2$ composite lattice (basal plane and overlayer) of cubic symmetry, both lattice systems are characterized by the *same* valence, namely, $\nu = 6$; we conjectured that our results might cast light on the efficiency of "docking" in $d = 2$ versus quasi-two-dimensional systems. Recall that in his analytic study of random walks on lattices of a given symmetry,

[1] The (very efficient) algorithm used to calculate the area under the piecewise continuous curves displayed in the figures was designed by Professor P. Hammer of Hollins College; her assistance is much appreciated.

having a single trap, and subject to periodic boundary conditions, Montroll [6] derived asymptotic expressions for the dependence of $\langle n \rangle$ on the system variables $[N, d, \nu]$. He proved that for $d = 2$ triangular lattices ($\nu = 6$), the mean walklength behaves as $\langle n \rangle \sim (\sqrt{3}/2\pi) N \ln N$. He also showed that for simple cubic lattices (also, $\nu = 6$), $\langle n \rangle \sim 1.52N$. Thus, for fixed $[N, d]$, one expects from Montroll's theory that (in the limit of large N) trapping will be more efficient in the higher dimension.

In the growth scenarios considered here, while the docking of a diffusing adatom occurs at a single site (trap), the stochastic event is complicated by the morphology of already deposited adatoms. Furthermore, the composite lattices considered here are quasi-two-dimensional, not "fully" $d = 3$ dimensional, so Montroll-type estimates are only approximate guides to the relative efficiency of the diffusion-reaction event. One can, however, use the calculated areal values for formation of platelets, thick dendrites, and thin dendrites to construct the following ratios:

$$\text{Platelet}\,(d > 2)/\text{platelet}\,(d = 2) = 0.869$$

$$\text{Thick dendrite}\,(d > 2)/\text{thick dendrite}\,(d = 2) = 0.750$$

$$\text{Thin dendrite}\,(d > 2)/\text{thin dendrite}\,(d = 2) = 0.830$$

For all three morphologies (platelet, thick dendrite, and thin dendrite), the overall efficiency of self-assembly on the basal plane is enhanced by enlarging the diffusion space to $d > 2$ (by including a single overlayer). However, as documented earlier, if one enlarges the diffusion space to include multiple overlayers, thereby simulating a $d = 3$ dimensional lattice, the diffusing adatom spends most of its time wandering around in the depletion zone before docking on the basal plane. This suggests that there is a spatial "bifurcation point", that is, a point where adding additional overlayers compromises, rather than enhances, the efficiency of the diffusion-reaction process. Simply stated, keeping fixed the number N of sites on the basal plane, while increasing the number n of overlayers so that the overall diffusion space has $n \times N$ sites, the advantages of expanding the diffusion space from $d = 2$, to $2 < \mathrm{d} < 3$, to $d = 3$ are (eventually) lost.

In light of the above results, it is of interest to recall the observations of Potter and Rossman [14], who note that dendrite formation occurs "within the rock matrix (internal), along fracture surfaces of the rock, or on its surface," and the experimental data reported by Chopard et al. [18] that show dendrites "emerging from cracks in limestone," or "within the plane of a crack inside a quartz crystal." It would appear that restricting the dimensionality of the diffusion space to "cracks" and "fractures" (where $2 < d < 3$) enhances the deposition of mineral dendrites.

Perhaps the most remarkable morphological difference between the dendrites formed in the two mineral systems considered here [17, 18] is the "perfect" symmetry of the zincite "snowflakes" formed in ZnO synthesis versus the fractal-like,

but not perfectly symmetric, branches of hollandite dendrites. In their comprehensive computational study of snow crystal growth, Gravner and Griffeath [30] state that their "algorithm assumes a mesoscopic (micrometer) scale of basic units for the ice crystal and water vapor, which eliminates inherent randomness in the diffusion and the attachment mechanism," noting that "by contrast, any three-dimensional approach based on microscopic dynamics is completely beyond the scope of present computing technology." Once a "building block" is introduced to "seed" the symmetry, Graver and Griffeath demonstrate that, implementing their algorithm, a manifold of pathways can lead to a manifold of final symmetric "snowflake" patterns.

Our results, both for $d = 2$ and for $2 < d < 3$, were obtained by introducing a single-site (structureless) "seed," documenting that the primary determinant influencing the evolution of patterns is geometrical. We show that, owing to steric constraints, different pathways to the same final structure are characterized by markedly different efficiencies. Whereas the final structures generated, faceted platelets (Figs. 1a and b and 2a and 2b) or fractal-like dendrites (Figs. 1c and 2c), are indeed symmetric (by design), intermediate stages in the evolution of the final structure certainly are not. It follows that, to the extent a deposition process is interrupted before the final stage of self-assembly is realized, it is highly improbable that the structures formed will have the remarkable symmetry of snow crystals.

Acknowledgments

The author wishes to thank Professor Harry Gray of the Beckman Institute, California Institute of Technology for his hospitality during the time this manuscript was in preparation. Particular thanks are given to Professor Kenneth G. Libbrecht, Norman Bridges Laboratory of Physics, and Professor George R. Rossman, Division of Geological and Planetary Sciences, at Caltech for their insights and suggestions on aspects of the work presented here. Continuing interaction with Professor Gregoire Nicolis of the University of Brussels and Dr. Catherine Nicolis of the Institut Royal Météorologique de Belgique on matters addressed in this chapter is deeply appreciated.

References

1. Y. Saito, *Statistical Physics of Crystal Growth*, World Scientific, Singapore, 1998.
2. (a) Y. W. Mo, J. Kleiner, M. B. Webb, and M. G. Lagally, *Phys. Rev. Lett.*, **66**, 1998 (1991). (b) Z. Y. Zhang, X. Chen, and M. G. Lagally, *Phys. Rev. Lett.*, **73**, 1829 (1994). (c) J. E. Vasek, Z. Y. Zhang, C. T. Salling, and M. G. Lagally, *Phys. Rev. B*, **51**, 17207 (1995). (d) A. Zhang and M. G. Lagally, *Science*, **276**, 377 (1997).
3. "Monte Carlo and Kinetic Monte Carlo Methods – a tutorial," in *Multiscale Simulation Methods in Molecular Sciences*, Lecture Notes, NIC series, J. Grotendorst, N. Attig, S. Blugel and D. Marx, Eds., ForshungszentrumJulich, 51–76 (2009).
4. J. J. Kozak, *Adv. Chem. Phys.*, **115**, 245 (2000).
5. (a) G. H. Weiss, *Aspects and Applications of the Random Walk*, North Holland, Amsterdam, 1994. (b) N. Wax, *Selected Papers on Noise and Stochastic Processes*, Dover Publications, New York,

1954. (c) C. W. Gardiner, *A Handbook of Stochastic Methods*, 2nd ed., Springer-Verlag, New York, 1985.

6. E. W. Montroll, *J. Math. Phys.*, **10**, 753 (1969).
7. R. Davidson and J. J. Kozak, *J. Phys. Chem.*, **102**, 7393 (1998).
8. R. Davidson and J. J. Kozak, *J. Phys. Chem.*, **102**, 7400 (1998).
9. P. A. Politowicz and J. J. Kozak, *Langmuir*, **1**, 429 (1985).
10. P. A. Politowicz and J. J. Kozak, *J. Phys. Chem.*, **94**, 7272 (1990).
11. K. G. Libbrecht, *Rep. Prog. Phys.*, **68**, 855 (2005).
12. J. J. Kozak, C. Nicolis, and G. Nicolis, *J. Chem. Phys.*, **126**, 154701 (2007).
13. J. J. Kozak and G. Nicolis, unpublished work.
14. R. M. Potter and G. R. Rossman, *Am. Mineral.*, **64**, 1219 (1979).
15. J. E. Post, *Proc. Natl. Acad. Sci. USA*, **96**, 3447 (1999).
16. D. A. McKeown and J. E. Post, *Am. Mineral.*, **86**, 701 (2001).
17. C. Li, G. Li, C. Shen, C. Hui, J. Tian, S. Du, Z. Zhang, and H.-J. Gao, *Nanoscale*, **2**, 2557–2560 (2010).
18. B. Chopard, H. J. Herrmann, and T. Vicsek, *Nature*, **353**, 409 (1991).
19. J. S. Langer, *Science*, **243** 1150 (1989).
20. T. Vicsek, *Fractal Growth Phenomena*, World Scientific, Singapore, 1989.
21. T. A. Whitten and L. M. Sander, *Phys. Rev. Lett.*, **47**, 1400 (1981).
22. R. Q. Hwang, J. Schroder, C. Gunther, and R. J. Behm, *Phys. Rev. Lett.*, **67**, 3279 (1991).
23. Th. Michely, M. Hohage, M. Bott, and G. Cosma, *Phys. Rev. Lett.*, **70**, 3943 (1993).
24. H. Roder, E. Hahn, H. Brune, J.-P. Bucher, K. Kern, *Nature*, **366**, 141 (1993).
25. H. Roder, K. Bromann, H. Breune, C. Borgano, and K. Kern, *Phys. Rev. Lett.*, **74**, 3217 (1995).
26. M. Hohage, et al., *Phys. Rev. Lett.*, **76**, 2366 (1996).
27. H. Brune, et al., *Surf. Sci.*, **349**, L115.
28. J. J. Kozak, G. Nicolis, and C. Nicolis, *J. Chem. Phys.*, **113**, 8168 (2000).
29. C. Nicolis, J. J. Kozak, and G. Nicolis, *J. Chem. Phys.*, **115**, 663 (2001).
30. J. Gravner and D. Griffeath, *Phys. Rev. E*, **79**, 011601 (2009).
31. G. R. Rossman,What are Manganese Dendrites?. Available at http://minerals.caltech.edu/FILES/DENDRITE/Index.html.

WHAT CAN MESOSCOPIC LEVEL *IN SITU* OBSERVATIONS TEACH US ABOUT KINETICS AND THERMODYNAMICS OF PROTEIN CRYSTALLIZATION?

MIKE SLEUTEL,[1] DOMINIQUE MAES,[1] AND ALEXANDER VAN DRIESSCHE[2]

[1]*Structural Biology Brussels, Vrije Universiteit Brussel, Brussel, Belgium*
[2]*Laboratorio de Estudios Cristalográficos, University of Granada, Granada, Spain*

CONTENTS

Advances in Chemical Physics, Volume 151: Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanoscale Materials, First Edition. Edited by Gregoire Nicolis and Dominique Maes.

I. INTRODUCTION

A. Protein Crystallization: An Intricate Part of Structure Determination

Protein crystallization has been a separate research area for over 30 years now. The major driving force for its development has been the strong growth in molecular biology, where the main aim is to understand biological systems from a microscopic perspective. This requires a multilevel approach where both the thermodynamics that govern the collective properties of a system of macromolecules need to be studied, as well as the molecular-level details of the constituting components. The latter requires insights into the protein structure/function relationship to fully reveal the inner workings of the molecular machinery. To reach this goal, structural biologists have benefited immensely from the determination of the three-dimensional structures of proteins using X-ray diffraction methods. The first structural entries in the protein databank based on this method date back to 1976 [1]. Since then, macromolecular structure determination at an atomic resolution has seen an enormous increase in impetus. The number of published structures has increased exponentially with a doubling time of approximately 3.8 years, reaching the 60,000 mark in 2010. Such vast efforts have been made to reach a single goal, that is, the vindication of the fundamental tenet of molecular biology, namely that true understanding of biological reactions is derived from the structure/function relationship of the participating proteins. When aspiring to achieve this goal, scientists are faced with multiple bottlenecks in the gene-to-structure *pipeline*; a term first coined in high-throughput structural genomics projects that seek to catalogue the protein structures associated with the whole genome of a target organism [2]. These bottlenecks include cloning, expression, "solubilization," purification, obtaining diffraction-quality crystals, phasing, structure solving, and relating structure to function. Macromolecular crystallization still poses one of the greatest challenges en route to determining the 3D structure and success depends strongly on the scientists' ability to obtain low-defect density crystals of high compositional and structural uniformity. A natural starting point to elucidate and overcome the difficulties in resolving this crystallization bottleneck is to

understand, through fundamental and applied research, the nature of the barriers associated with the three main processes of crystallization, that is, macromolecular interaction, nucleation, and growth.

B. Protein Crystallization *In Vivo*

Protein crystals are not only the focus of interest from the *practical* standpoint of structure determination, they are used by nature as well [3]. This might seem contradictory given that millions of years of "evolutionary negative design" are believed to have stripped proteins of their propensity to self-assemble into aggregates or crystals [4]. Given that cellular solutions are crowded multicomponent protein mixtures with packing fractions of up to 40%, avoiding self-aggregation would seem quite logical. Under these cramped conditions, undesirable crystallization would most likely compromise the cells' viability, serving as a strong negative selection mechanism. Indeed, many *condensation* diseases are related to the loss of protein solution stability resulting in the formation of a condensed phase, such as crystals, polymers, dense liquids, aggregates, etc. For example, liquid–liquid demixing of γ-crystallins in the eye retina underlies the pathology of cataract formation. Here, lens opacification temperatures are strongly dependent on specific surface amino acid residues [5–7]. A similar sensitivity is found for hemoglobin, where point mutations that lead to intraerythrocytic crystals [8] or polymers [9] contribute to the pathogenesis of sickle cell anemia [10]. However, many examples do exist of proteins that do crystallize *in vivo*. These crystals serve specific purposes, such as storage [11], secretion [12], solid-state catalysis [13], encapsulation [14], sealing [15], and so on. One of the most well-known examples is the regulated secretion of crystalline granules of insulin [16]. Additionally, protein crystals are also being used in biomedical areas of research where a slow sustained release of administered medications *in vivo* is being pursued. To achieve such a maintained release over prolonged periods of time, control is required over both the number and size-distribution of the crystallized substances, such as insulin for the case of diabetes [17].

C. Model System for Self-Assembly in General

Protein crystallization has benefited greatly from 100 years of research on the crystallization of small molecules. In fact, most models on both protein nucleation and growth are firmly rooted within the vast theoretical frameworks that have been developed for small molecule crystallization. Although it is generally accepted that macromolecular and small molecule crystallization are both governed by the same underlying principles, there is still great merit in studying protein crystallization. In addition to their role in structural biology and biomedicine, protein crystals can be used as a model system for crystallization and first-order phase transitions in

general. Protein crystallization is a process of self-assembly by molecular (self)-recognition at a high degree of reliability. Many functions of biological molecules depend on the correct recognition of complementary partners, and as such are of fundamental interest. Hence, insights into the intricacies of the recognition mechanisms involved in protein crystallization may provide a new and/or improved perception of self-assembly processes in general. In comparison to colloids, thermal equilibration times remain relatively short for proteins due to the smaller molecular mass, making macromolecules a more *meaningful* model system [18]. For small molecules, however, crystal growth rates are within two to three orders of magnitude higher than protein crystals [19]. Such fast incorporation kinetics complicates the detailed observation of sequential discrete growth events. Conversely, for proteins the interval time between two successive molecular attachment events and the molecular sizes are compatible with the temporal and spatial resolution of current observation techniques. For instance, recent experimental and theoretical studies on (protein) nucleation (reviewed in Ref. 20 and Chapter 4 of this book; [21]) have lead to novel insights and modifications of existing models that have already been applied successfully to different crystallization systems.

In this chapter, however, we focus on postnucleation events, that is, crystal growth. In the last two decades, experimental protein crystal growth studies have been focused on growth phenomena operating on either a microscopic or macroscopic scale. For excellent overviews on the matter, the reader is encouraged to examine the following Refs. 18, 22–30. The most widely used techniques have been atomic force microscopy and phase shifting Michelson interferometry, respectively. The former has allowed the investigation of nanoscale processes with molecular-level resolution on a (highly) local scale. These include the identification and characterization of dominating growth mechanisms at various supersaturation intervals, visualization of the formation of critical nuclei, molecular level defect formation such as point defects, vacancies, stacking faults, dislocations, impurity incorporation, etc. Michelson interferometry, however, has been successful at elucidating growth phenomena that operate on a crystal-wide scale that do not necessarily require superior lateral resolution. Main subjects using this technique have been normal growth and surface morphology, the coupling of incorporation kinetics and mass transport, step bunching and accompanying growth instabilities, and so on. Processes that operate at the intermediate level and thus require the combination of a crystal-scale observation area and the ability to discern elementary steps on the surface have remained largely unexplored. This can be attributed due to limited accessibility to such regimes by the techniques present at the time. *In situ* analysis of these mesoscopic processes is, however, vital in linking elementary incorporation processes at the kink-level to macroscopic crystal properties such as crystal face kinetics, diffraction characteristics, crystal morphology, and so on. This mesoscopic level has become fully accessible to the experimentalist with the advent of laser confocal differential interference contrast microscopy.

In the current contribution, we provide a nonexhaustive overview of the recent advances in mesoscopic protein crystal growth studies. We focus on a number of case studies to demonstrate the strengths and weaknesses of the mesoscopic experimental approach. We begin by discussing the three main techniques for studying crystal growth (Section II). In the following three sections, crystal growth close to equilibrium (Section III), at intermediate driving forces (Section IV), and at high supersaturation levels (Section V) is addressed, respectively. Next, effects of impurity species on crystal growth kinetics is focused on (Section VI) followed by concluding remarks (Section VII).

II. TECHNIQUES FOR MESOSCALE *IN SITU* OBSERVATIONS OF PROTEIN CRYSTAL GROWTH

During the last decades, protein crystal growth from aqueous solution has been monitored *in situ* using different observation techniques such as optical microscopy [31], atomic force microscopy (AFM) [26, 32], regular and phase shifting Michelson interferometry [33, 34] (PSMI), and the more recently developed laser confocal microscopy combined with differential interference contrast microscopy (LCM-DIM) [35]. With the exception of regular optical microscopy, the remaining techniques can be used for the collection of accurate data on mesoscopic dynamic processes occurring on crystal surfaces. AFM [26, 32] and LCM-DIM [35–38] can image *in situ* spiral hillocks and two-dimensional (2D) islands, allowing the direct measurement of tangential step velocities and 2D nucleation rates. Using interferometry, it is very difficult to achieve a resolution that allows the observation of monolayer steps on spiral hillocks or 2D islands of elementary height [39]. Hence, measurements of step velocities, v_{step} using interferometry are performed on spiral hillocks [33, 34] and step velocities are obtained indirectly from the slope p of a hillock and the normal growth rate R by [40]

$$R = pv_{\text{step}} \tag{1}$$

where the slope p of the vicinal is given by the ratio of the step height h and the interstep distance λ.

An appropriate choice of the observation technique is vital for the collection of useful data with the frequency and lengthscale resolution best suited for the problem at hand. Inappropriate usage of experimentally determined kinetic data can lead to biased results of limited applicability. Therefore, in this first section, we evaluate the usefulness of AFM, PSMI, and the novel LCM-DIM as observation techniques at a mesoscopic level for crystal growth kinetics. Results on growth kinetics obtained by these techniques are compared and differences are discussed, pointing out the advantages and limitations thereof.

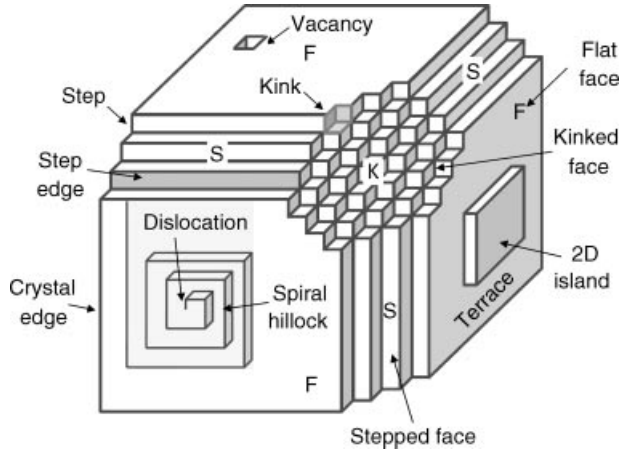


Figure 1. Kossel crystal composed of kinked (K), stepped (S), and flat (F) faces.

A. F-Faces

The mechanism of crystal growth is unambiguously determined by the structure of the crystal surfaces. Crystal habits are composed of three types of faces [41, 42], kinked (K), stepped (S), and flat or facetted (F) faces (Fig. 1). The former two have the highest density of uncompensated bonds and thus the lowest density of molecules. They are rough on a molecular scale and abundant in favorable growth sites (kinks) and therefore grow very fast. Owing to the substantial kink density in K faces and to a lesser extent in S faces, it follows that during growth these 2 types of faces disappear first, giving rise to a crystal fully enclosed by low index F faces [43]. The structure of these flat F faces [42] of perfect crystals, however, offer no kink sites and rely on kink generating mechanisms to procure their growth. The high energetic costs and kinetic barriers for the creation of kinks and/or steps result in relatively slow growth rates and a flat surface topography. From an experimentalists point of view, these characteristics make F-faces highly compatible with both the temporal and spatial resolution of the techniques discussed below. Unlike K and S faces, which have a limited lifetime due to their morphology, F faces can grow indefinitely making them ideal research objects for fundamental crystal growth studies. Not surprisingly, all crystal growth studies (including this work) have been focused on growth phenomena on F-faces.

B. Measuring Growth Kinetics with AFM, LCM-DIM, and PSMI

Atomic Force Microscopy A typical AFM image of a growing (110) face of a tetragonal lysozyme crystal is shown in Fig. 2a. At relatively high

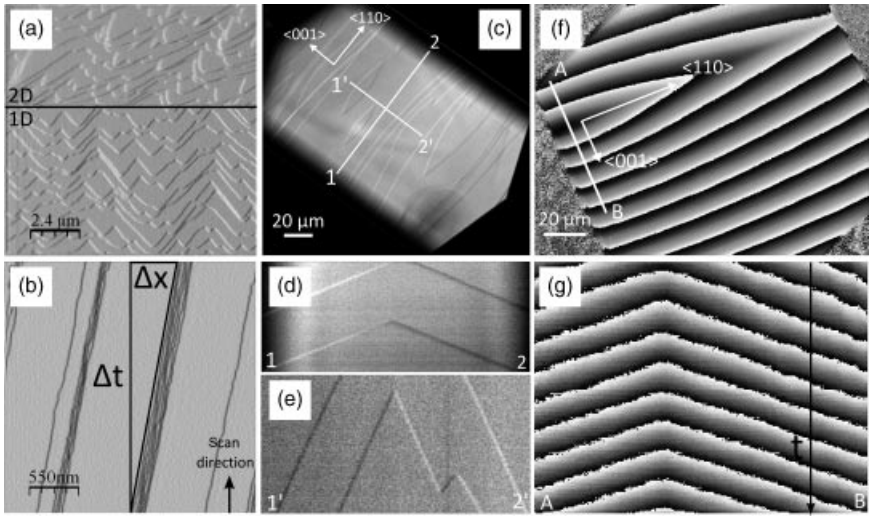


Figure 2. AFM, LCM-DIM, and PSMI images of growing (110) faces of tetragonal lysozyme. (**a**) AFM image showing nucleation and growth of 2D islands (black line indicates transition from a time–space scan to a space–space scan). (**b**) Time–space AFM scan of step trains advancing at constant velocity (the vertical represents time). (**c**) LCM-DIM image of an entire (110) face showing monolayer 2D islands. (**d** and **e**) Time–space plots of step advancement in the ⟨001⟩ and ⟨110⟩ directions. (**f**) A typical phase shift interferogram of a spiral hillock on a (110) face. Along the line A–B a time–space plot is constructed for the entire length of the experiment (**g**).

supersaturation ($C-C_e = 54.6\,\mathrm{mg\,mL^{-1}}$) crystal growth proceeds by 2D nucleation. The upper part (labeled "2D") shows the XY (space–space) image depicting a high density of 2D islands, which correspond to the relatively high supersaturation at which the crystal was growing. The lower part of the image (labeled "1D") shows a time–space image recorded with the slow-scan axis Y disabled so that the vertical direction encodes time [44]. This picture summarizes the kind of kinetic information one can obtain from an AFM experiment: 2D nucleation density, roughness of the surface, and step velocity. Figure 2b shows a time–space scan of steps on the (110) face growing at low supersaturation ($C-C_e = 6.9\,\mathrm{mg\,mL^{-1}}$) in the ⟨001⟩ direction. Step trains are moving at constant velocity and the interstep distance remains constant over time. Since the scan frequency is known, step velocity can be directly determined from these time–space images provided that the angle between the step line and the scan direction is known [45, 46].

Step velocities as a function of supersaturation measured using tapping (intermittent contact between AFM tip and surface) and contact mode (continuous contact) in the ⟨001⟩ direction are shown in Fig. 2a. Contact mode

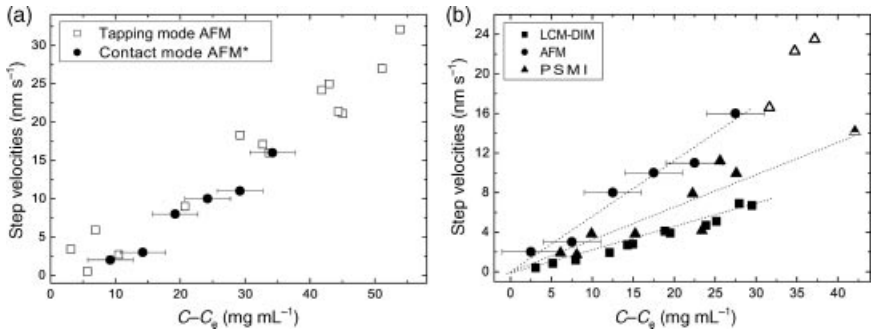


Figure 3. Step velocities measured in the ⟨001⟩ direction on the (110) face of tetragonal lysozyme crystals growing from purified lysozyme solution. (**a**) Comparison of step velocities measured by tapping and contact mode AFM. (**b**) Comparison of step velocities obtained by AFM, LCM-DIM, and PSMI. Open triangles indicate step velocities calculated from PSMI data obtained above a critical supersaturation ($C-C_e > 30\,\mathrm{mg\,mL^{-1}}$) where 2D nucleation becomes appreciable (*unpublished data from Nakada et al.).

AFM data are taken from Nakada et al. (unpublished data) and were collected under identical experimental conditions. No significant difference is observed between data obtained by these two AFM operational modes. The comparison of step velocities measured as a function of supersaturation using PSMI, contact mode AFM, and LCM-DIM in the ⟨001⟩ direction on (110) faces of tetragonal lysozyme are shown in Fig. 3b. As could be expected from the indirect nature of the measurement, the PSMI data show the largest scatter, but, more importantly, significant differences are observed between step velocities measured using different techniques. From the plot shown in Fig. 3b, it is clear that step velocities measured by AFM (either in tapping or contact mode) are significantly higher than those measured by LCM-DIM and PSMI. A similar discrepancy was observed for the (011) face of orthorhombic glucose isomerase (GI) crystals [38, 47].

The possible influence of AFM imaging on liquid dynamics and incorporation kinetics is still a matter of debate. With this technique, a cantilever is oscillating at a high frequency inside the solution (tapping mode) and scanning rapidly in a horizontal plane (tapping and contact mode), close to the crystal surface. Effects of cantilever movement during AFM observations were reported by Land et al. [48] who found that the growth rates of canavalin crystals were locally enhanced by the scan of the cantilever. Gliko et al. [49] hypothesized that large variations in step velocities of lumazine synthase crystals could be the result of the cantilever oscillation. Van Driessche et al. argue that cantilever movement disturbs the diffusive boundary layer

developed in the vicinity of a growing crystal, thereby exposing the step edge to a higher effective concentration [50]. Conversely, Gasperino et al. conclude from finite element modeling that mass transfer limitations between the liquid and crystal surface can lead to local solute depletion, thus reducing the surface supersaturation [51]. Alternatively, Chen et al. have reported absence of inhibition or acceleration of step propagation by cantilever scanning [52]. Given the wide spectrum of observations, additional comparative kinetic experiments on well-characterized model systems are required to unravel this matter further.

Laser Confocal Microscopy A representative LCM-DIM image of a growing (110) face of a tetragonal lysozyme crystal is shown in Fig. 2c. The vertical resolution of the image allows for the observation of individual steps on the surface. The same information available from AFM experiments can therefore also be obtained with LCM-DIM using the similar time–space images for the tangential growth of 2D islands or dislocation hillocks (Fig. 2d and e). These time–space images are constructed by extracting the pixels of a specific line in the bitmap (white lines in Fig. 2c) along which the step advancement over time is observed. Each of these lines from consecutive frames will be a row in the time–space picture. Figure 2d shows one such time–space plot for the birth and spread of a 2D island ($C-C_e = 9.9\,\mathrm{mg\,mL^{-1}}$). The island spreads with equal velocity, $v_{step} = 8.7\,\mathrm{nm\,s^{-1}}$, into the positive and negative ⟨110⟩ direction. The same observations were made for the ⟨001⟩ direction having steps growing at $v_{step} = 1.4\,\mathrm{nm\,s^{-1}}$ (Fig. 2e). Step velocities were found to be constant over time.

Normal growth rates can also be measured using LCM-DIM (if previous data about the step height is available), and AFM as well, simply by counting the newly formed layers over time, but this is only possible for supersaturation values low enough as to avoid step overlapping at a lateral scale smaller than the microscope resolution. In both cases the unambiguous identification of single monolayer steps is central to their use in the evaluation of kinetic data.

Phase Shifting Michelson Interferometry A typical PSMI interferogram of a spiral hillock is shown in Fig. 2f. From a series of phase shifting interferograms collected at regular intervals, a time–space picture along a certain line can be constructed (Fig. 2g) from which the normal growth rate can be directly determined by plotting the height at the given point against the corresponding time. Step velocities are then computed from the normal growth rate using Eq. (1) and the hillock slopes measured in the ⟨110⟩ and ⟨001⟩ directions on the space–space image. From such a time–space picture, step velocities can be directly obtained but the lateral resolution is much lower than the vertical resolution and data obtained in this way will have a much

larger error range. Step velocities obtained from PMSI data show significant scattering but are comparable to those directly measured by LCM-DIM at low supersaturation. For larger values ($\ln(C/C_e) > 1.3$), a steep increase in the step velocities is observed from the PSMI data (Fig. 2b). This is, most likely, the result of significant 2D nucleation on the terraces of spiral hillocks at higher supersaturations. This leads to an underestimation of p and subsequent overestimations of v_{step}. For mixed or purely 2D nucleation growth regimes, the slope p in Eq. (1) can no longer be approximated by h/λ. Hence, step velocities should not be readily extracted from PMSI interferograms without prior knowledge of 2D nucleation rates. Additional care must be taken when using PSMI to determine the slope of a hillock over a short lengthscale. Bunched steps that traverse the surface will temporally modify the slope of the hillock. When applying Eq. (1), these local slope changes can lead to artificial step velocity fluctuations. Thus, temporal and/or local slope variations explain a larger dispersion of the data points found for PMSI measurements compared to AFM or LCM-DIM.

With interferometry, obtaining data at low supersaturation is difficult because normal growth rates are hardly detectable in the case of 2D nucleation-mediated growth, while lateral (step) growth rates are large enough to be detected. For example, 2D heterogeneous nucleation at low supersaturation has never been observed with interferometry for protein crystal growth (and seldom for small molecules) [53, 54], while AFM studies were able to detect heterogeneous nucleation for thaumatin [55, 56] and catalase [56] crystal growth. With LCM-DIM, 2D nucleation could be studied in detail for tetragonal lysozyme crystals [37] and orthorhombic glucose isomerase crystals [38].

The significant differences in step velocities found in the growth rates of lysozyme are reflected back in the kinetic proportionality constant between the supersaturation and the tangential growth rate, that is, the step kinetic coefficient β_{step} defined by

$$v_{step} = \beta_{step}\Omega\,(C - C_e) \tag{2}$$

where Ω is the crystal volume per molecule and C and C_e the bulk and equilibrium protein concentration. For the data shown in Fig. 3, one obtains for β_{step} $4.4 \times 10^{-5}\,\mathrm{cm\,s^{-1}}$, $2.6 \times 10^{-5}\,\mathrm{cm\,s^{-1}}$, and $1.7 \times 10^{-5}\,\mathrm{cm\,s^{-1}}$ for AFM, LCM-DIM, and PSMI, respectively. β_{step} is, however, an intrinsic kinetic parameter that should be technique-independent. The scatter in β_{step} therefore partly reflects the experimental artifacts, introduced by the different techniques that cloud an unambiguous physical interpretation. Potential artifacts are improper AFM settings resulting in high tip–surface interaction forces, step advancement rates comparable to or larger than scanning

tip velocity, step velocity extraction of 2D hillocks from PMSI interferograms with nonzero 2D nucleation rates, and so on. Differences in sample cell geometry may also have a profound influence on the recorded data. Sample cell thickness and bulk solvent accessibility to the crystal surface will significantly affect mass transport rates in the system. Edge effects, crystal size distribution, and local depletion by neighboring crystals complicate matters further, greatly influencing local surface supersaturation. Of course, one can introduce forced convection by imposing a solutal flows and thus diminishing gradients in the system and exposing the surface to bulk solute concentrations. However, potential enhancement of impurity effects on step kinetics and induced step bunching should be taken into account. These effects certainly contribute to the observed scatter and make the determination of β_{step} a nontrivial task. In general, AFM data will correspond better to situations were forced convection is present (i.e., no concentration gradient around the growing crystal) while LCM-DIM data, and PSMI data (after suitable correction) will correspond better to diffusive mass transport situations or reduced natural convection situations (i.e., when a developed concentration depletion zone exists around the crystal).

C. Spatial and Temporal Resolution

Unambiguous quantitative information on the surface topography is available from AFM and PSMI, but not from LCM-DIM. Therefore, additional information from the other techniques is needed to compute normal growth rates from LCM-DIM data. The vertical resolution of LCM-DIM is close to 1 nm and with AFM a vertical resolution around 1 Å is possible [26, 57]. Thus, both techniques have adequate vertical resolution for studying single-step dynamics on protein crystals (usually nanometer step height). The theoretical vertical resolution of the PSMI is around 1 nm, but in practice this resolution depends on the reflectivity of the crystal face, the disturbances originating at the solution or the cover glass, and the overall quality of the optical parts. Consequently, this resolution is worse for the protein surface having low reflectivity. Due to the limited resolving power of interferometry for direct observation of single steps, experimental data must be carefully evaluated to avoid biased interpretations or extrapolations beyond the resolution of the instrument. Although, recently developed, advanced PSMI techniques extend the vertical resolution to the molecular scale and are able to observe single steps at low supersaturation [39].

Lateral resolution of PSMI and LCM-DIM is limited to that of the microscope lens used for imaging, and can be estimated using Abbe's equation (0.61λ/numerical aperture). For a $40\times$ lens the upper limit is approximately $0.7\ \mu$m. Therefore, observation of processes below this scale requires the use of AFM. This fact is clearly seen in the space–time pictures used to evaluate the

step velocity (Fig. 2b as compared to Fig. 2d and e). In the case of AFM, this figure shows a certain "roughness" not observed in LCM-DIM images because it exists at lengthscales smaller than the lateral resolution of the microscope. If one is interested in studying the kink dynamics, AFM is superior, but when studying mesoscopic step dynamics, the LCM-DIM is preferred.

For a fixed amount of information, the lateral resolution is the inverse of the field of view, so high lateral resolutions means small field of view. LCM-DIM and PSMI allows the observation of large areas, while AFM is more limited in lateral field of view, which can be a problem when dislocation hillocks and 2D nucleation operate simultaneously on the same surface or when the growth rate shows inhomogeneities over the crystal face. The maximum scan area for most AFMs is approximately $120 \times 120\,\mu m^2$. This limitation is principally due to the operational set up of the AFM. Figure 2a and c show typical AFM and LCM-DIM images of a (110) face growing by the 2D nucleation mechanism. The AFM image was scanned in 256 s and covers an area of $10 \times 10\,\mu m$, while with LCM-DIM 9.6 s were necessary to scan an area of $800 \times 800\,\mu m$.

D. Optimal Operational Range

The selection of an experimental technique is not just a matter of fitting the instrument to the mass transport and physical model relevant to the problem at hand. Because the three techniques have different vertical, lateral, and time resolutions, the length- and timescales of the problem is relevant as well. PSMI shows the highest data acquisition rate but the lowest vertical and lateral resolution while AFM has the lowest acquisition rate and the highest resolution. LCM-DIM fits in between, having a vertical resolution comparable to that of AFM (at least for protein crystal growth problems), a lateral resolution such as the one of PSMI and an intermediate data acquisition rate. Therefore, AFM is the observation technique of choice for slow, nanoscale growth processes, LCM-DIM is a microscale observation technique with nanoscale vertical resolution useful for slow/medium growth processes and PSMI is the best-suited technique for fast growth processes requiring only micrometric spatial resolution. Consequently, the three techniques complement each other for studies on crystal growth kinetics.

In the case of protein crystal growth, the supersaturation range for which normal growth rates can be directly determined using LCM-DIM and even more so for AFM, is outside the range used in most crystallization experiments for the production of crystals for X-ray diffraction. At such low growth rates a crystal will take several months or even years to grow to a reasonable size. AFM observations are most reliable and meaningful at low supersaturation levels where, for example, kinetics at step edges can be observed *in situ*. With LCM-DIM excellent observations can be carried out in the low to middle supersaturation range but at high supersaturation levels the lateral resolution is not high enough to distinguish

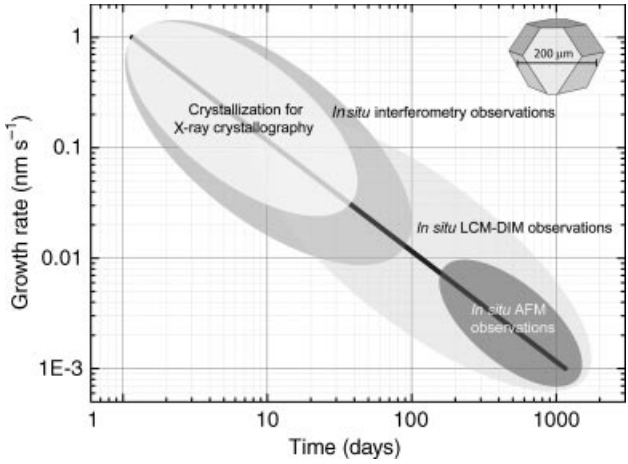


Figure 4. The typical operational supersaturation range for each technique is shown compared to the typical supersaturation ranges in which crystals for X-ray crystallography are nucleated and grown. The line represents crystals with a size of 200 μm. (Modified after P. Dold and K. Tsukamoto, unpublished.)

individual steps of 2D islands. The use of interferometry is most appropriate in the middle to high supersaturation range while at low supersaturation levels macroscopic growth processes are very slow and measurements become sensible to large errors. A schematic representation of the typical operational supersaturation range for each technique is shown in Fig. 4.

III. PROTEIN CRYSTAL GROWTH CLOSE TO EQUILIBRIUM

Protein crystal growth at low supersaturation levels can be studied by most conventional techniques on a macro-, meso-, and nanoscopic level. It is therefore a highly documented subject in literature. The very slow advancement (nm s^{-1}; Å s^{-1}) of the edges of unfinished layers (steps) on the surface makes this regime an excellent candidate for in depth, molecular resolution investigations. These works include molecular-level defect formation, impurity effects, 2D lattice characterization, layer generation mechanisms, etc on a wide variety of protein systems, ranging from bacterial to eukaryotic and viral proteins. However, most works center either on the quantitative aspects at a mesoscopic level or on the qualitative characteristics at a nanoscopic level. Here we focus on a combination of both, that is, nanoscopic observations of dynamic events at the kink level.

First we discuss the determination of the enthalpy and entropy of crystallization by monitoring step kinetics of lysozyme and glucose isomerase at near-equilibrium

conditions using LCM-DIM. Then we use AFM to study the step dynamics of triosephosphate isomerase from which we infer information on the solute-to-crystal pathway and the elementary growth units that incorporate into the kinks. This protein has a low step edge free energy combined with a low-step kinetic coefficient (unpublished data), resulting in very high step densities (beyond the LCM-DIM limit) that advance very slowly, that is, ideal for AFM.

A. Enthalpy and Entropy of Crystallization

A practical advantage of observing step dynamics close to equilibrium, that is, growth and dissolution, is that the solubility of proteins in aqueous solution can be precisely determined. From these data the numerical values for enthalpies and entropies of crystallization can be calculated with significantly reduced experimental errors compared to other macroscopic methods. As mentioned above, with LCM-DIM elementary steps can be readily observed, and at the same time an entire crystal surface can be visualized. This allows for a detailed observation of the surface step dynamics over time. With the help of an accurate temperature control system, equilibrium temperature intervals can be determined for a wide concentration range as a function of temperature for globular proteins. From the obtained solubility data, enthalpy, entropy, and Gibbs free energy of crystallization can be calculated.

1. Solubility Curve Determination

The LCM-DIM method is based on the *in situ* observation of single steps on a protein crystal surface, which allows a fast and precise determination of solubility as a function of temperature. When the temperature of a protein solution is set below/above the equilibrium temperature, growth/dissolution of the crystal surface will occur for normal solubility dependence and vice versa for retrograde solubility. Using this technique, the equilibrium temperature is defined as the average of the interval between the minimum temperature at which dissolution is observed and the maximum temperature at which growth is observed in a certain time period. Dissolution of a crystal surface can be recognized by either retreatment of single steps or the crystal edge and/or the nucleation of etch pits, depending on the level of undersaturation. Growth is characterized by step advancement and 2D nucleation. The LCM-DIM setup not only allows us to observe *in situ* single steps on a protein crystal surface but also the crystal edges can be clearly visualized and thus growth or dissolution of a crystal surface can be distinguished. LCM-DIM requires only a slight change (micrometer range) in the position of the steps or the crystal edge to detect growth or dissolution of a crystal surface. By establishing the limits of growth and dissolution as a function of temperature, the true equilibrium temperature is located in this temperature interval. It should be mentioned that

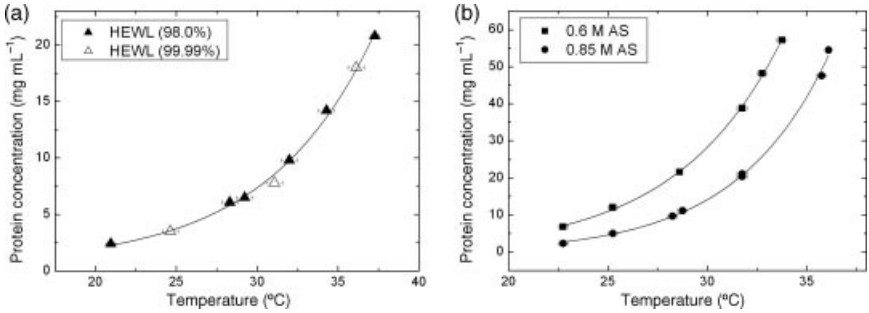


Figure 5. (**a**) The solubility of tetragonal lysozyme in 50 mg mL^{-1} NaCl. Solid triangles indicate commercial grade lysozyme (≥98.0%, Worthington) and open triangles indicate highly purified lysozyme (99.99%, Maruwa Food Inc.). (**b**) The solubility of glucose isomerase in 0.6 M (squares) and 0.85 M (circles) ammonium sulfate. The curves are guides for the eye.

the accuracy of this method is only limited by the stability of the protein solution and the precision of the temperature control system. With this method, it is also possible to establish solubility curves as a function of precipitant concentration, type, ionic strength, and so on.

The temperature solubility curves for GI and HEWL were determined with great precision (±0.5°C) by *in situ* observation of the crystal surface. The solubility curves for one salt concentration in the case of HEWL (Fig. 5a) and for two salt concentrations for GI (Fig. 5b) are shown. For each protein concentration, the temperature interval is plotted in between which the equilibrium temperature is located. For both proteins the solubility increases with temperature, as is the case for most globular proteins [58]. Also note that the solubility decreases with increasing precipitant concentration.

2. *Determination of Enthalpy and Entropy of Crystallization*

The solubility data shown in Fig. 5 were used to characterize the thermodynamics of crystallization using the van't Hoff equation [59]

$$\ln K_e = -\left(\frac{\Delta H^o - T\Delta S^o}{RT}\right) = \frac{\Delta G^o}{RT} \tag{3}$$

where K_e is the equilibrium constant for crystallization, T is absolute temperature, R is the universal gas constant, and ΔG°, ΔH°, ΔS° the standard free energy, enthalpy, and entropy of crystallization, respectively. If we assume solution ideality (i.e., the activity coefficient γ_e is equal to unity) the crystallization equilibrium

constant K_e can be represented as

$$K_e = a_e{}^{-1} = \left(\gamma_e \frac{C_e}{C^o}\right) \approx (C_e/C^o)^{-1} \tag{4}$$

where C_e is the solubility and $C^o = 1$ mol kg^{-1} is the concentration of the solution in the standard state. In the case of protein solutions, using this value for the standard state is thermodynamically not self-consistent [60]. But for the sake of comparison, this value was used because most thermodynamic values reported in literature are obtained using this value. The selection of a different standard state (e.g., 1 mmol kg^{-1}) does not affect the value of ΔH° and the shift in determined values of ΔS° and ΔG° are relatively minor and will therefore not affect the conclusions about the underlying physical processes [18]. Combining Eqs. (3) and (4) yield the well-known equation

$$\ln\left(\frac{C_e}{C^0}\right) = \frac{\Delta H^o}{RT} - \frac{\Delta S^o}{R} \tag{5}$$

From the linear plot ($\ln(C_e/C^\circ)$ versus $1/T$), one obtains for the slope and intercept $\Delta H^\circ/R$ and $-\Delta S^\circ/R$, respectively (Fig. 6a). Note that the linear relationship of the data indicates that the enthalpy of the system is temperature independent in the studied temperature range. The numerical values for enthalpies and entropies of crystallization are summarized in Table I. Using Eq. (3), the standard free energy can also be determined (Fig. 6b).

The negative entropy contribution for both proteins indicates entropy loss during crystallization, and hence, the entropy change disfavors crystallization. This entropy loss is due to the constrained translational and rotational degrees of freedom of the protein molecules and the release, trapping, and rearrangement of water

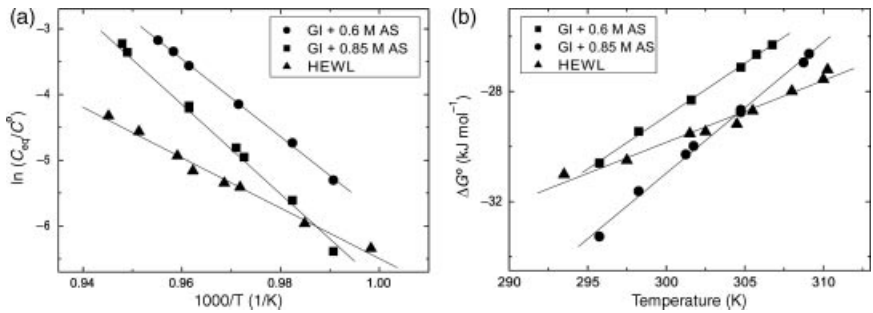


Figure 6. (**a**) Log-linear plot of the solubility data for GI and HEWL used to obtain the enthalpies and entropies of crystallization. The straight lines are least squares fits to the van't Hoff equation; (**b**) Gibbs free energy of crystallization as a function of temperature for GI and HEWL.

TABLE I
Thermodynamics of Crystallization of Glucose Isomerase and Hen Egg White Lysozyme

Precipitant	ΔH (kJ mol^{-1})	ΔS (J mol^{-1} K^{-1})
0.6 M ammonium sulfate: GI[a]	-144 ± 1	-370 ± 2
0.85 M ammonium sulfate: GI[a]	-174 ± 1	-462 ± 4
0.91 M ammonium sulfate: GI [61]	-160 ± 40	-420 ± 100
50 mg mL^{-1} NaCl: HEWL[a]	-96 ± 1	-220 ± 4
25 mg mL^{-1} NaCl: HEWL [62]	-110	-241
25 mg mL^{-1} NaCl: HEWL [63]	-72	-216

[a]This study.

upon attachment to the crystal [18]. This reduces the magnitude of the crystallization free energy from the values set by the enthalpy and in this way leads to higher solubility. For crystallization to occur this entropy contribution needs to be compensated. For the studied temperature range the entropic effect is compensated by the negative enthalpic change, and the crystallization process is exothermic. For GI, there is a significant difference in ΔH° and ΔS° between 0.6 M and 0.85 M AS, indicative of the salting-out effect at higher salt concentrations. In Table I the enthalpies and entropy values of crystallization of GI and HEWL reported in this study and by other groups for similar solution conditions (i.e., pH, precipitant) are shown. No significant differences are found but the margin of error is strongly reduced for the values obtained by the LCM-DIM method.

B. Step Dynamics

The formation of new layers and their subsequent advancement by attachment events slows down significantly as the system reaches equilibrium. This leads to an effectively arrested state where the surface is kinetically *frozen* when observed macroscopically. On a molecular and mesoscopic scale, however, steps meander around an equilibrium state due to the interplay between step stiffness and molecular dynamics, which include attachment and detachment from the step due to exchange with the bulk reservoir and surrounding terraces, line diffusion and kink formation due to thermal motion of the molecules. The slow kinetics and molecular lengthscales at which these dynamics operate make AFM a particularly insightful technique to unravel the underlying molecular-level mechanisms that govern crystal growth.

For example, evidence for the prevailing incorporation mechanism can be sought in the study of step dynamics [64]. Step incorporation/detachment events operate on a timescale ($\sim$s) that is two orders of magnitude smaller than a typical AFM image acquisition time ($\sim$200 s). This incompatibility of timescales can be remedied by disabling the slow scan axis direction, hereby scanning the same line

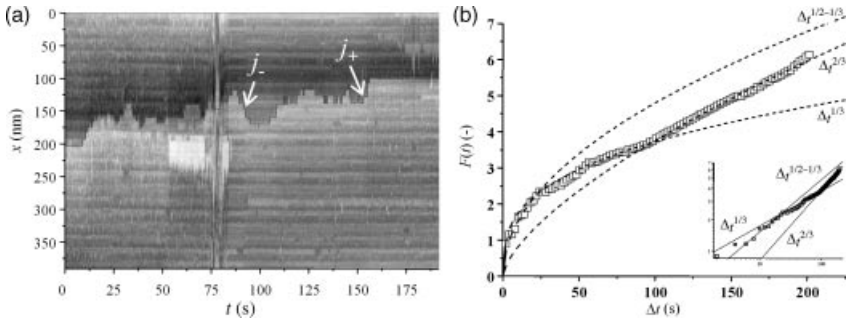


Figure 7. (**a**) Time evolution of the step position shows the molecular step displacement with attachment events (+) and detachment events (j_-) changing the step coordinate by one unit. (**b**) Time correlation curve, characterizing mean square displacement of step location as a function of time Δt.

perpendicular to the nominal direction of the step edge as a function of time (1D sampling). Moreover, this technique allows us to separate the influence of spatial fluctuations on step dynamics from temporal fluctuations [65]. An example of such a step trace image for the case of triosephosphate isomerase from *Thermotoga maritima* (TmTIM) is shown in Fig. 7a. These dynamical step fluctuations can be the result of various transport phenomena. One is where mass transport occurs only along the step through a process called true line diffusion. Another is where step perturbations result from mass exchange of the step with the solid surface phase (2D surface diffusion) and the surrounding liquid phase (direct incorporation). The presence of line diffusion can be identified by interpreting the time correlation curve $F(\Delta t)$ [66] of the spatial displacements of the step edge using equilibrium step dynamics models [64, 67]. The obtained correlation curve displays three different time regimes. At smaller timescales the correlation curve can be fitted using a power law with either exponent $1/2$ or $1/3$, at intermediate timescales $1/3$ and at larger timescales the exponent $2/3$ is obtained. As pointed out by Ihle et al. [64], the predicted sequence of exponents at intermediate and large timescales depends on the limiting transport mechanism and the presence of a weak or strong Schwoebel effect. If supply is limited to diffusion, transport on the terrace or in the bulk solution is slow. Conversely, in the kinetically limited case slow attachment/detachment events are the rate-determining step. Yet, irrespective of the limiting transport mechanism, step perturbations brought about by a conformational reorganization of the step through line diffusion are predicted to follow a power law $1/4$. The absence of a $1/4$ power law in the fitted data presented in Fig. 7b indicates that line diffusion contributes only marginally to the temporal step fluctuations. This allows us to attribute the step trace displacement events to the exchange of growth units with the surrounding phases, that is, the terrace and solution phase.

C. Elementary Acts of Crystal Growth

The frequencies of attachment j_+ of building units to kinks and their detachment j_- are elementary parameters of crystal growth and are used in almost all models that describe step and face advancement. Despite their fundamental importance, they remain highly elusive parameters to attain experimentally. As such, quantitative data on j_+ and j_- remains very scarce in literature [52, 68, 69]. Typically, only the microscopic step growth rate is followed experimentally. Through the knowledge of the interkink distance one can calculate the average net rate of growth unit addition, that is, $j_+ - j_-$. To measure j_+ and j_- experimentally, however, molecular resolution AFM experiments are required to resolve the individual attachment/detachment spatially (Fig. 8a). To obtain adequate temporal resolution as well, one-dimensional time traces need to be recorded. From such 1D-sampled images, the one-dimensional size distribution of units attaching to and detaching from the step edge can be determined. For example, for the case of TmTIM, the 7 nm peak can be interpreted as the addition of a single tetramer when compared to the X-ray structural data shown in Fig. 8b. Likewise, the small 15 nm peak can be correlated with the attachment/detachment event of a single unit cell comprising three tetramers. It is conceivable that the dominating peak at 11 nm stems

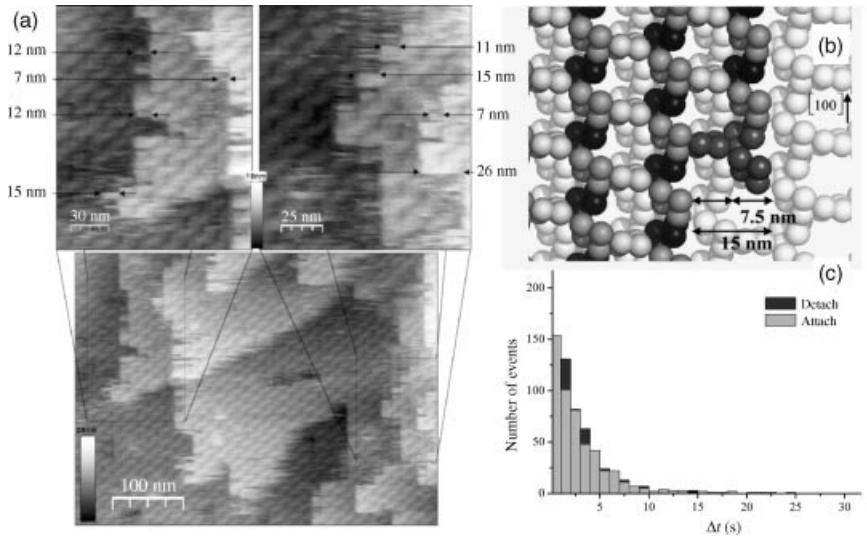


Figure 8. (**a**) Molecular resolution AFM height image of the (011) face of a TIM crystal. Thin horizontal lines at the step edges are the resultant of attachment/detachment events occurring on a timescale smaller than the temporal resolution of the system. (**b**) A [100] step edge on the (011) face. The unit cell is comprised of three tetramers (single sphere represents a single monomer). (**c**) Histogram of time intervals Δt between individual attachment (light gray) and detachment events (dark gray).

from an intermediate kinetic state, that is, a partial unit cell that is in the process of being completed and formed during the dead zone time of the experiment, that is, the temporal resolution of 0.4 s. The same quantitative result is obtained through normal mode imaging at a molecular resolution of the crystal surface. Here, we observe thin horizontal lines at the step edges that are the resultant of attachment/detachment events occurring on a timescale smaller than the temporal resolution of the system (Fig. 8c). Then what is the growth unit for the crystallization of TmTIM? Most likely the dominating growth unit is a single TmTIM tetramer. Although larger sized attachment/detachment events are observed, it is probable that they are formed on a timescale smaller than 0.4 s. Indeed, when looking at the kinetics of incorporation, more specifically the time intervals between individual attachment and detachment events, we find an exponential decay as a function of time. From this data, we conclude that the time interval between a large part of the events is of the order of 1 s or smaller (Fig. 8c). Clearly, the kinetics of the system operates on a timescale equal to or smaller than the temporal resolution obtained in our experimental setup. This is exemplified by the absence of a clear maximum in the time distribution of attachment events. Such a maximum would be expected because molecules incorporate at finite velocities in the step. This suggests that the addition of units with linear dimensions larger than a single TmTIM tetramer, can be the result of a series of incorporation events within the timeframe of 0.4 s. To get a better view at the molecular dynamics in the kink, a high-speed AFM should be used on these types of systems.

D. Growth Unit Pathways Toward the Kink

Under conditions of a nonvanishing activation barrier for nucleation and kink formation (i.e., below the kinetic roughening transition [70] and with moderate molecular thermal energy [71], otherwise said below the thermal roughening temperature) F-type interfaces are smooth surfaces where the relatively rare incorporation sites are situated along the molecular rows that define the edges of the unfinished advancing crystalline layers traversing the surface. For F faces where kinks are typically widely spaced, an interesting competition between mass transfer through bulk and surface diffusion arises. With the presence of kink-free terraces in between the (linear) sinks, that is, steps, the growth units can now adhere to the surface, diffuse two-dimensionally and incorporate into kinks (Fig. 9); a pathway previously not possible on the S and K faces that are only present in the early stages of crystal growth. The dominating pathway (either bulk diffusion in the convection-free liquid boundary layer normal to the surface or 2D surface diffusion of adsorbed molecules (admolecules) on terraces in between the steps [72]) can have a profound effect on the surface morphological stability with respect to local perturbations [24]. This can be understood in the following way. Irrespective of the most prevalent solute pathway toward the kink, steps are fed by growth

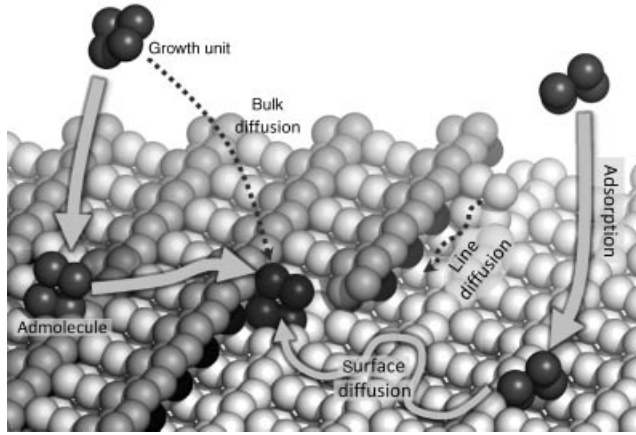


Figure 9. Two possible pathways for a growth unit from solution into a step: direct incorporation from solution; step integration as a result of surface diffusion. Reorganization of the step occurs through line diffusion.

units that are diffusive toward the incorporation sites. When the incorporation of molecules at the step proceeds faster than the rate at which diffusion can supply new growth units, a local depletion of solute molecules is established. The zone of lower solute concentration around a step can be considered to be a half cylinder centered at the step. If the "supply fields" of two or more steps overlap, the local solute concentration decreases even more, leading to a mutual deceleration of the steps involved. Such an effective *step–step attraction* can lead to significant step train instabilities (step bunching) eventually leading to the formation of macrosteps. The latter are regions of local high step densities that incorporate impurities at a different rate than solitary steps. This causes impurity gradients within the crystal bulk and subsequent stress build-up that can lead to striations, mosaicity and cracks, which result in a diminished crystal quality. Now, scaling arguments suggest [73] that competition for supply from the bulk should be too weak to cause any discernible effects on step advancement. Step bunching as a consequence of diffusion field overlap should therefore only be expected when steps are solely being fed by admolecules. Not surprisingly, determining the molecular pathway that a molecule follows from the solution to the F-interface of the crystal bulk is vital to understand the mechanisms of growth. Regardless, the significance of surface diffusion in crystallization from solution remains a long-standing problem of fundamental interest [25].

Although the total change in the average energy of a particle as it moves from the solution to the interface is the same for both pathways, the energy activation barriers the particle has to overcome can differ for both trajectories. The relative

heights of these barriers that need to be surmounted, will determine the dominating transport mechanism. Mesoscale evidence that favors a multistage process involving adsorption on the surface and two-dimensional diffusion toward the steps has been presented for a few model proteins only, that is, canavalin (AFM, [45]), insulin (AFM, [74]), and the (110) face of lysozyme (PSMI, [75, 76]). In contrast, for the (010) face of orthorhombic lysozyme it has been shown that surface diffusion has a negligible contribution to tangential growth [77]. One of the possible caveats of interpreting mesoscale data in favor of direct incorporation is that the requirement of interstep interaction, that is, surface diffusion field overlap, is not met in the tested experimental conditions. This problem does not occur when studying step dynamics at a molecular level using AFM where interstep interaction is not a prerequisite.

Earlier we showed that line diffusion has a negligible influence on the step dynamics of TmTIM. Now we can test the existing models of direct incorporation from solution and surface diffusion by examining the ratio of elementary attachment j_+ and detachment frequencies j_- of single growth units. For the ratio of the fluxes of molecules entering and exiting the steps, models predict [78, 79]

$$\frac{j_+}{j_-} = \frac{C}{C_\mathrm{e}} \qquad \text{(direct incorporation)} \tag{6}$$

$$\frac{j_+}{j_-} = \frac{n_s}{n_\mathrm{e}} \qquad \text{(surface diffusion)} \tag{7}$$

with C and C_e the actual and equilibrium bulk concentration, and n_s and n_e the actual and equilibrium surface concentration, respectively.

For the time trace [80] shown in Fig. 7a, we have $j_+/j_- \leq 1.2$. This ratio is an upper estimate because any attachment/detachment events that occurred within the temporal resolution remain undetected. The growth velocity for that step edge is $0.62\,\mathrm{nm\,s^{-1}}$, which is very low. For the C_e we have $0.043\,\mathrm{mg\,mL^{-1}}$, giving for $C/C_\mathrm{e} = 4.7 > j_+/j_-$. This inequality indicates that the model of direct incorporation from solution does not amply predict the measured statistics of exchange in the kink. This suggests that bulk diffusion alone cannot be considered the sole supplier of growth units and that the surface diffusion pathway should also be considered. Let us test the model of surface diffusion. If we assume that all molecules entering the kink are admolecules, then according to Eq. (2) we have $n_\mathrm{s}/n_\mathrm{e} = j_+/j_-$. Given that n_s and n_e are not readily attainable experimentally (for exploratory works on this matter, the reader is referred to Sazaki et al.) we can approximate their ratio assuming Langmuir adsorption: $n_\mathrm{s} = n_\mathrm{sat} C/(C+B)^{-1}$ where n_sat is the maximum density of admolecules and B the Langmuir constant. Assuming B is positive (i.e., n_s increases for increasing C) we get the following for a nonsaturated state $n_\mathrm{s}/n_\mathrm{e} \leq C/C_\mathrm{e}$ or equivalently $j_+/j_- \leq C/C_\mathrm{e}$, which is confirmed by our data. The

mechanism of incorporation of admolecules thus corroborates our observed lower ratios of influx to outflux.

Given that protein surface diffusion has been estimated to be 3–4 order of magnitude slower [81] than diffusion in the bulk, how can admolecules still contribute appreciably to step kinetics/dynamics? One of the potential reasons is that macromolecules experience a high rotational entropic barrier for kink incorporation due to their size, geometrical complexity, and structural anisotropy [82]. Because no sufficient torque reorienting the biomolecule for correct incorporation is expected, the molecule approaches the kink at random orientation [27]. The probability that a molecule will arrive at a kink with an orientation that is relevant for regular attachment is estimated to be 10^{-2}–10^{-3} [78]. In addition to this configurational obstacle, macrobond patches between future neighbors (on the incoming molecule, the molecules belonging to the underlying layer, and the molecules in the kink) should partially shed their hydration layer. As opposed to direct incorporation from solution where all these events need to operate in a concerted way and therefore result in a large activation barrier, surface diffusion is a two-step process where the barriers for adsorption and incorporation into the step are separated [45]. Quantitative data on adsorption and incorporation barriers of ferritin ($U_{\text{adsorption}} \leq 26\,\text{kJ mol}^{-1}$; $U_{\text{step}} \leq 44\,\text{kJ mol}^{-1}$) [52] compared to the potential barrier for direct incorporation $\sim$88 kJ mol^{-1} strengthen this notion. Additionally, confining the search space of the diffusing species from the bulk liquid phase to a two-dimensional surface may also increase the efficiency of the growth process. For instance, Adam and Delbrück showed that biological diffusion processes such as membrane-associated reactions and bacterial chemotaxis may benefit from a reduction in dimensionality. Elaborating on this notion, Berg and Purcell found that the mass transfer efficiency of 2D diffusion depends on the average distance between sinks on the surface, with *sinks* being kinks in our case, or more generally steps. It is not inconceivable that a similar mechanism may operate for the crystallization case. We also note that dominance of a single pathway may be regime dependent, that is, supersaturation dependent. This concept was touched upon in the work of Garza-Lopéz et al. who approached the problem of postnucleation self-assembly using a lattice-statistical approach to define regimes where either mechanism will dominate based on the ratio of accompanying kinetic rate constants [83].

IV. PROTEIN CRYSTAL GROWTH AT INTERMEDIATE DRIVING FORCES

Up to now, growth under intermediate driving forces has been the most investigated supersaturation range by AFM and interferometry. From these observations, pioneering works by Durbin and Feher [151] revealed for the first time the growth mechanisms of protein crystals. After this breakthrough, many studies followed,

mainly focusing on understanding the basic principles underlying these mechanisms of growth. By measuring interfacial growth kinetics, such as advancement rates of steps emanating from spiral dislocations or 2D nucleation rates on terraces, fundamental parameters such as kinetic coefficients and surface free energy could be determined. Studies on impurity effects and defect generation revealed that all protein crystals deviate strongly from the idealized perfect case. These processes take place at the mesoscopic level and are the ideal playground for LCM-DIM. In this chapter, we therefore zoom-in on crystal growth phenomena in the intermediate supersaturation range, which require a crystal-scale observation area combined with medium lateral resolution and the ability to discern elementary steps. First we discuss polyhedral (in)stability issues arising from a critical crystal size and the associated stabilizing and destabilizing factors. Then we focus on a more exotic mode of layer generation, previously undocumented for protein crystal growth, that is, (long term) repeated 2D-nucleation.

A. Polyhedral Stability-Limit: Starvation Flaws

The interplay between surface growth processes and the distribution of concentration around a growing crystal determines the shape and stability of the crystal habit. When the kinetics of protein attachment onto a growing crystal face removes protein more rapidly from the region adjacent to the crystal face then mass transport can replenish the supply of growth units, a significant concentration gradient is established. In the absence of relatively fast convective currents, an isotropic protein depletion zone is formed which creates local solutal conditions that can differ greatly from the bulk composition. The accompanied slow-down of the crystal growth process is considered to promote the growth of high quality crystals [84]. By removing convection, spatial irregularities and temporal oscillations in solute transport that cause the formation of defective regions in the crystal are reduced. As such, a more stable mode of growth is attained [73, 85, 86]. However, the effect of ablation of convection is system dependent and may either be advantageous or disadvantageous to the crystal quality. For instance, it has been shown that a reduction of the mass transport rate can enhance defect-causing step density and velocity fluctuations [87] or augment impurity uptake [82, 88–90]. Such a mass transport regime that is governed by diffusion can be reached by working in a zero- or microgravity environment [91–93], in gelled solutions [94–96], or by lowering the ratio of the buoyancy and viscous forces of the system (capillary [97, 98], closely spaced glass plates [62, 99–101], high molecular weight PEG's [102]). There is one other case where elimination of convection will not be beneficial for protein crystallization and ultimately crystal quality, that is, loss of polyhedral stability [40, 103]. It is a kinetic phenomenon where, due to the specific interplay of surface kinetics and mass transfer, the crystal is no longer capable to retain its polyhedral shape and large depressions develop on the habit faces. Since the

destabilizing mechanisms scale-up with crystal size and the limiting of nutrient supply, polyhedral stability is expected to occur in the same conditions required for the formation of a depletion zone. The geometry of this metastable halo encircling the crystal does not coincide with the polyhedral shape of the crystal. Consequently, edges and vertices of growing crystals are better supplied with nutrients than facet centers. This surface concentration inconstancy is referred to as the Berg effect [104] and can have a pronounced effect on surface kinetics and ultimately surface morphology [105]. Polyhedral instability is a long standing problem in the crystallization of small molecules due to their fast incorporation kinetics (for example, see Refs. 106–109) and it is important for applications where large protein crystals are required, for example, neutron diffraction. Initial reports of polyhedral instability of protein crystals were made by Nanev et al. [110, 111] for ferritin, lysozyme, and trypsin. Further steps toward a deeper understanding were made by Vivares et al. who studied the concentration gradients in the vicinity of glucose isomerase crystals using confocal scanning fluorescence microscopy [102]. They achieved ablation of convection through the use of a high molecular weight PEG (10 kDa) and correlated the polyhedral instability of crystals to the presence of a depletion zone.

However, to gain deeper insights into the underlying principles that govern crystal stability, it is essential that one has detailed knowledge on the crystal height profiles, types and rates of layer generation, and step velocity distribution profiles across the crystal surface. Since the typical lengthscale of the Berg effect is commensurable with crystal size, classic surface imaging techniques such as AFM are less ideal due to the limited surface scanning area. It is therefore recommended to employ the less invasive optical method LCM-DIM which, given the fast scan rates (on average 5 s/image), large scanning area (millimeter range) and medium lateral resolution ($\pm 0.6\ \mu$m) is ideal to probe mesoscopic kinetic surface phenomena [35, 112]. Here we discuss two aspects of polyhedral (in)stability [113]: (1) the main extrinsic instability mechanism that perturbs the system, that is, the Berg effect and (2) the intrinsic anisotropy of the face kinetic coefficient that stabilizes the face against bulk solvent induced perturbations. We will demonstrate that crystals larger than a critical size will inevitably lose their polyhedral stability due to the limits of the stability mechanisms. When this critical point is reached, the faceted form is no longer capable to remain similar to itself and a central starvation flaw will form.

1. Instability Mechanism: The Berg Effect

Figure 10a shows the manifestations of the Berg effect on 2D nucleation kinetics and step density for protein crystallization in a nongelled medium. A clear increase in step density is observed from the crystal edge (left-hand side) toward the facet center (right-hand side). The reverse trend is visible with regard to the presence

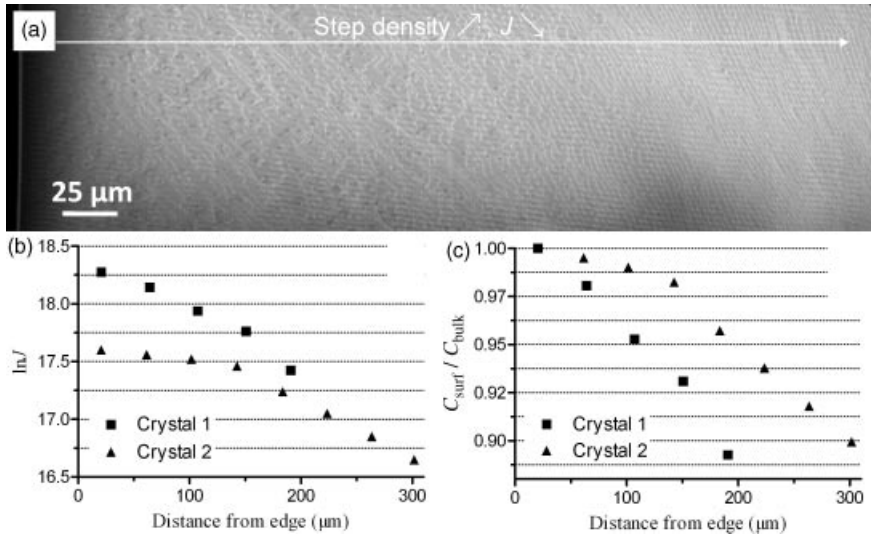


Figure 10. (**a**) Step and 2D island density increase and decrease, respectively, as a function of distance to the crystal edge (left-hand side) as a consequence of the Berg effect, that is, nonconstant surface concentration. Nucleation rates (ln$J$) (**b**) and relative depletion $C_{\text{surf}}/C_{\text{bulk}}$ (**c**) as a function of surface coordinate along the white arrow in (a).

of 2D islands, that is, the surface density of 2D islands decreases as a function of distance from the crystal's edge. A similar deceleration of kinetics is observed for the lateral step velocity v_{step}.

More quantitatively, time-lapse imaging of the crystal surface allows to determine 2D nucleation rates J ($\text{m}^{-2}\text{s}^{-1}$) as a function of the distance from the edge x (in µm, taken to be zero at the crystal edge). Figure 10b shows an ln$J$ versus distance from the edge plot obtained for two different crystals. A (kinked) linear decrease in nucleation rates is observed as a function of surface coordinate. The kink in the plot for crystal 2 can be attributed to the geometry of the system, that is, the angle and distance between the crystal surface and the reactor walls and edges. Given the high nucleation rates, the prevailing step-generating mechanism is homogeneous 2D nucleation (HON) [38] characterized by a specific edge free energy of $\kappa_{\text{HON}} = 6.0 \pm 0.3 \times 10^{-13}\,\text{J m}^{-1}$. Through a model for the steady-state 2D nucleation rates J derived from classical nucleation theory [43]

$$\ln J = \ln\left(\omega^* \Gamma Z\right) - \frac{\pi \kappa_{\text{HON}}^2 s}{k^2 T^2 \ln\left(C_{\text{surf}}/C_{\text{e}}\right)} \tag{8}$$

where ω^* is the frequency of attachment of molecules to the critical 2D nucleus, Γ is the Zeldovich factor, Z is the steady-state admolecule surface concentration, s is

the surface area of a single molecule in the critical 2D nucleus (9.7×10^{-17} m^2) and C_{surf} and C_e are the protein surface and equilibrium concentration, respectively, the local surface supersaturation $\sigma_{surf} = \ln(C_{surf}/C_e)$ can be calculated. Such an approach requires that the observed retardation in surface kinetics is a direct consequence of surface concentration gradients and not of step diffusion field overlap. This requirement is fulfilled here because AFM experiments [38] have shown that no step–step interaction occurs for distances >75 nm (3–4 lattice parameters), well below the minimum interstep distances in this work. When no significant temperature gradients are assumed, the relative surface depletion C_{surf}/C_{bulk} can be determined, where C_{bulk} is the protein bulk concentration, which we equate with C_{edge}. In reality, most likely $C_{edge} \leq C_{bulk}$, however, we disregard this inequality since we focus here on the determination of relative surface concentration gradients rather than absolute gradients. This leads to the C_{surf}/C_{bulk} versus distance from edge plots in Fig. 10c. A depletion of approximately 11% is measured at $\pm 250\,\mu$m from the crystal edge. Although this does change the microprofile of the crystal surface, the overall face's shape remains identical, that is, macroscopically flat. We can conclude that in the presence of relatively small surface concentration gradients (moderate Berg effect), shape stabilizing factors [114] (e.g., anisotropy of face kinetic constant, surface tension and capillarity, nonuniform distribution of growth-retarding impurities, temperature gradients) succeed in preserving the polyhedral shape of the crystal. However, shape-destabilizing factors (e.g., nonuniformity of (impurity) concentration distribution) scale-up with crystal size and can lead to morphological defects. In the following section, we discuss the various (micro)structures of the surface in response to supersaturation inhomogeneity while retaining a flat macroscopic interface.

Figure 11a shows one of the "flattest" and most stable modes of growth, that is, at random homogeneous multilayer 2D nucleation (note that layer-by-layer growth would be flatter) with an average slope of zero. The random nature of this process (average island density is constant as a function of surface coordinate) indicates the absence of any significant concentration gradients on the surface. Any local perturbations in surface height show no temporal stability and fade away. When the crystal size increases or supply becomes hampered by reactor edge effects, concentration gradients arise that affect both the kinetics and localization of 2D nucleation. Figure 11b illustrates this phenomenon: step generation becomes limited to specific local areas of higher surface concentration leading to the formation of a 2D hillock. In response to the concentration gradients, the step density increases as a function of distance to the nucleation center until eventually the interstep distance becomes smaller than the critical length of a 2D nucleus or the supersaturation becomes smaller than the critical supersaturation for 2D nucleation. This effect enhances the "confinement" of 2D nucleation and hence step generation to a specific region. While Fig. 11b shows straight step trains, Fig. 11c shows loss of step straightness at a specific distance from the crystal's edge. This step concavity is the

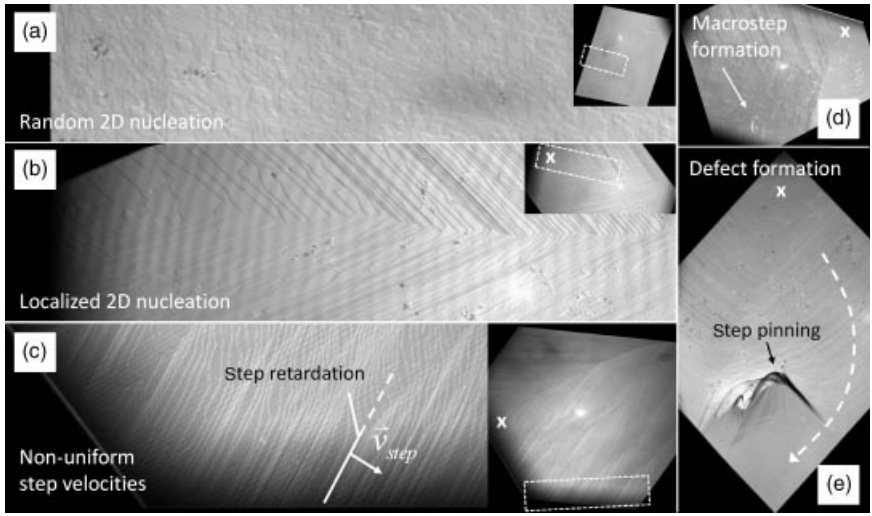


Figure 11. The violation of the morphological stability of a flat, crystallographic low index face (011) leads to the appearance of a central depressed area: (**a**) step generation through 2D nucleation occurs at random across the entire surface; (**b**) surface concentration gradients lead to preferential 2D nucleation at crystal edges, leading to step trains crossing the vicinal surface; (**c**) step train velocity becomes nonconstant along the average direction of the step resulting in increased step concavity toward the center of the crystal face; (**d**) loss of stability of semiequidistant step trains leads to the coalescing of steps and formation of macrosteps downhill of the nucleation center; (**e**) step pinning combined with increasing diminishment of step velocity in the face center lead to a singular critical point where quasizero step velocity is reached, creating a depressed valley downhill of the critical point, dashed white line indicates direction of step advancement. Dashed white boxes in zoom-out pictures indicate zoom-in area (a–c). White *x*'s denote preferential 2D nucleation centers (b–e).

result of a nonconstant step velocity along the average direction of the steps. Since this effect is observed for all the face's edges at specific edge distances, the concavity can be attributed to surface concentration gradients and rule out other step retarding possibilities such as impurities. Next, very high step densities giving rise to diffusion-field overlap and impurity pinning in the face's center can eventually lead to loss of stability of the step trains. This gives rise to the step bunching and subsequently formation of macrosteps present in the center of the face (Fig. 11d). Macrosteps can then lead to an additional slowing down of step advancement in the central areas of the crystal face. All these effects combined, finally result in the appearance of a singular point/region on the surface where quasizero step velocity is reached. Upon reaching this point a cascade of events enhances the instability even more, that is, step density increases further and faster by the arrival of new steps to the near-zero-velocity region and step advancement slows down due to step–step interaction. Ultimately, this nonmoving macrostep (shock wave) leads

to the formation of a central expanding depression (Fig. 11e). This microscopic transformation is the onset of loss of polyhedral stability. It is important to keep in mind that this is a kinetic phenomenon that can be remedied by decreasing the supersaturation or increasing the material supply through (forced) convection.

2. *Stability Mechanism: Face Kinetic Coefficient Anisotropy*

For a moderate Berg effect the crystal remains able to retain its polyhedral facetted shape through the available compensation mechanisms. However, in the absence of convective currents, the concentration difference between edge and center $C_{\text{edge}} - C_{\text{center}}$ scale-up with [40] $\beta_f R/D$, where β_f is the face kinetic constant, R is the crystal size, and D is the diffusivity, and eventually the crystal starts to change its habit, turning into a skeleton or dendrite. Here, we discuss the most general (albeit limited) factor maintaining the regular polyhedral shape of a growing crystal [115], that is, the anisotropy of face growth rate $V(\boldsymbol{n})$ as a function of orientation $\boldsymbol{n}$. Figure 12a shows the microcompensation profile in response to supersaturation inconstancy, that is, a gradual increase in step density toward the crystal center leading to a concave surface profile. The local slope p increases from 0.004° to 0.09°; the reverse trend is witnessed for the tangential step velocity v_{step}, that is, steps slow down farther from the edge (Fig. 12b). The crystal can offset this step retardation generated by the supersaturation inconstancy by increasing

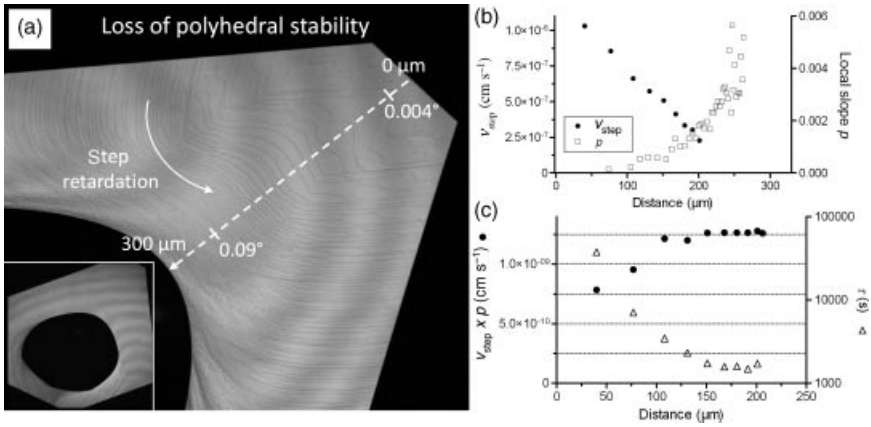


Figure 12. (**a**) The crystal's microscopic compensation profile in response to supersaturation inconstancy is limited by the maximum of the kinetic coefficient β. A clear failing to preserve polyhedral stability results in the formation of a large central circular depression, (inset) zoom-out of the region presented in (a); (**b**) tangential step velocity v_{step} (•) and local slope p (□) as a function of distance from the edge in the direction indicated by the dashed white arrow in (a); (**c**) $v_{\text{step}} \times p$ (•) and terrace exposure time τ (Δ) as a function of edge distance.

its local slope p. The obtained surface curvature generates a higher kinetic coefficient in the central area compared to the face periphery and is able to counteract the destabilizing effect of supersaturation surface gradients. This is illustrated in Fig. 12c, where $v_{\text{step}} \times p$, which we use as a measure of the normal growth rate, remains constant over the first $\pm 200\,\mu$m along the trajectory indicated by the dashed white arrow in Fig. 12a. Note the smaller values of $v_{\text{step}} \times p$ close to the edge. In this region, where 2D nucleation still occurs, the face growth rate cannot be approximated with $v_{\text{step}} \times p$ and the normal growth rate is underestimated.

For regions closer to the center an "avalanche like loss of stability" sets in. The increase in $\beta_f R/D$ as a result of the compensation profile leads to an additional rise in $C_{\text{edge}} - C_{\text{center}}$ thereby increasing the demand for an increased surface curvature. Eventually, an abrupt decline is reached in the anisotropy of the kinetic constant (expected for larger p) and the kinetic coefficient $b(p)$ attains its maximum ($\partial b/\partial p = 0$). $\partial b/\partial p$ can be calculated using the following relationships [40]:

$$p_{\text{center}} - p_{\text{edge}} = \frac{1}{\Theta} \frac{\sigma_{\text{edge}} - \sigma_{\text{center}}}{\sigma_{\text{edge}}} \tag{9}$$

$$\Theta = \frac{1}{b}(\partial b/\partial p) \quad \text{with} \quad b(p) = \beta(p)\sqrt{1+p^2} \tag{10}$$

where Θ is the anisotropy of the face kinetic coefficient $\beta(p)$. We derive the local slope $p(x)$ from the surface profile (Fig. 13a) and calculate $\sigma(x)$ from the step velocity assuming no diffusion-field overlap using [24] $v_{\text{step}} = \Omega\beta_{\text{step}}\ (C - C_e)$

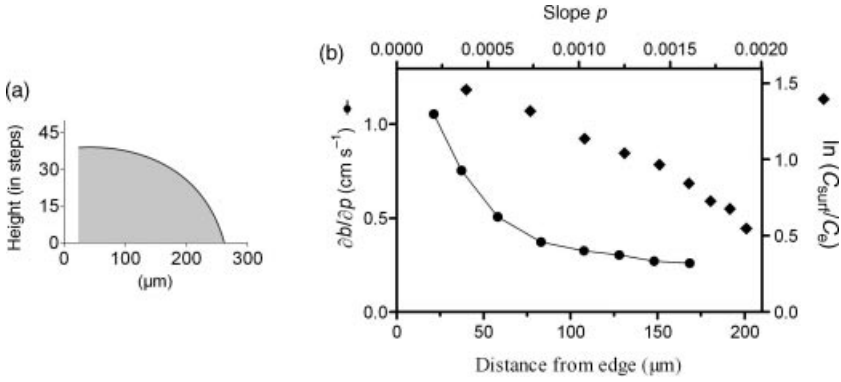


Figure 13. (**a**) Surface relief of the compensation profile in Fig. 12a; (**b**) surface supersaturation derived from step velocities as a function of edge distance, bottom and right x- and y-axis, respectively, (♦); anisotropy of the kinetic coefficient as a function of local slope p, upper and left x- and y-axis, respectively, (•).

with $\beta_{\text{step}} = 4 \times 10^{-4}\,\text{cm}\,\text{s}^{-1}$. The results between 0 and 200 μm from the edge are summarized in Fig. 13b.

The decrease in surface supersaturation is far more pronounced than in Fig. 10b. Such significant depletion can arise when the crystal face is in close proximity to the reactor wall, rendering the face's center poorly supplied and enlarging the Berg effect [113]. We show here for protein crystallization using LCM-DIM, that both $\sigma_{\text{surf}} = \ln(C_{\text{surf}}/C_{\text{e}})$ and $\partial b/\partial p$ decrease steadily, demonstrating the diminishing strength of the compensation mechanism at higher slopes. For low step densities, the kinetic coefficient increases strongly with increasing local slopes. At larger slopes, however, the kinetic coefficient of the face becomes practically independent of the orientation. These data demonstrate that crystals larger than a critical size will inevitably lose their polyhedral stability and a central starvation flaw will form. When this critical point is reached, the faceted form is no longer capable to remain similar to itself. In practice the depression is offset from the center of the face and step generation is limited to only a single vertex as opposed to multiple vertices. We believe this to be the result of a nonzero angle between the average crystal face orientation and the reactor wall.

B. 2D Hillocks

As mentioned earlier, step bunching can have a profound effect on both the surface topography of the crystal face as well as the uniformity of the crystal bulk. Of course, for step bunching to occur, large step trains first need to be formed on the surface. Local perturbations, for example, protein, precipitant, and temperature gradients, can locally accelerate/decelerate steps thereby disturbing the step train. Instability mechanisms, such as competition for supply [116, 117], solutal flows [118, 119], impurity effects [120–122], stochastics of layer generation [123, 124], Ehrlich–Schwoebel effect [125–128], and the coupling of bulk transport with nonlinear interface kinetics [87] can then stabilize and even amplify these bunches into macrosteps as they move across the surface. Typically, the stable source of steps is assumed to be one or more spiral dislocations. However, large step trains can also be generated by the other main layer generation mechanism, that is, 2D nucleation. Typically, surfaces that grow in the supersaturation range where steps are generated by 2D nucleation, are relatively flat, that is, no significant slopes exist and step trains are not present. This is because 2D nucleation is a random mechanism where new layers are formed arbitrarily across the surface. New islands merge with steps already present on the surface thereby annihilating any significant height changes on the surface. If, however, the formation of new steps is limited to one specific region on the crystal surface, steps will expand from that area thereby forming a train of nonequidistant steps. This results in the formation of a "2D-hillock," which observed from a microscopic point of view (e.g., low resolution Michelson interferometry), may be mistakenly interpreted as a spiral hillock.

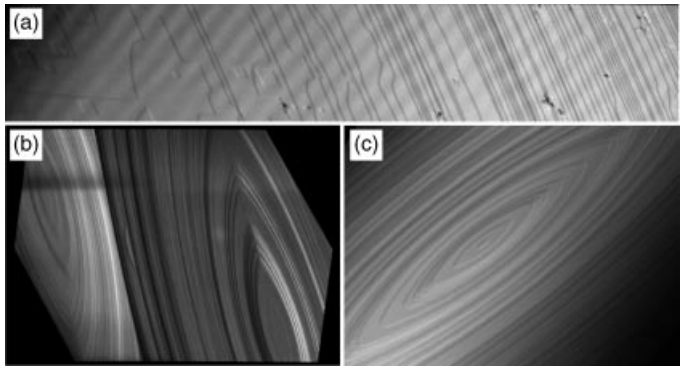


Figure 14. (**a**) 2D hillock on the (011) face of an orthorhombic glucose isomerase crystal; (**b**) mesoscopic image of the (110) face of a tetragonal HEWL crystal dominated by two 2D hillocks formed by 2D nucleation hotspots; (**c**) zoom-in area of a 2D nucleation hotspot.

To properly study such a system, there are two main requirements: (1) the imaging area needs to allow for visualization of the regions of step generation and of step train formation. Typically, the relevant lengthscale is commensurable to crystal size, ranging from 50 to 1000 μm; (2) adequate lateral resolution is required to discern between 2D nucleation and a spiral hillock. Fulfilling these requirements is the technique LCM-DIM. In Fig. 14, two scenarios are presented for the formation of a 2D hillock. The most straightforward case is where 2D nucleation is limited to the edges or apexes of the crystal due to the Berg effect, that is, significant surface concentration gradients. If the supersaturation in one apex is just high enough to allow for 2D nucleation and no spiral dislocations are present, this will yield a 2D hillock that covers the entire surface. A second, more exotic scenario has been observed to date for one protein only, that is, HEWL. Here, 2D nucleation is not confined to one region but is limited to quasione point, a nucleation "hotspot" of preferential island formation. In such a scenario, nucleation is no longer at random spatially, but does remain at random temporally. This creates an infinite supply of concentric islands (much akin to spiral hillocks) that are spaced irregularly (stemming from the erratic nature of the process) thereby resulting in a 2D hillock (Fig. 14b and c). Most likely, this is a case of heterogeneous 2D nucleation, that is, a local point of low step edge free energy or diminished nucleation activation barrier facilitated by, for example, a stacking fault [129] or fibrous material [130]. The stress generated by either a planar defect or a foreign particle incorporated into the lattice would locally increase the solubility at the hotspot. Subsequent slow dissolving should then generate a point-bottomed etch pit. This was, however, not observed for this case, indicating that further studies are necessary to clarify this rather special case of 2D nucleation.

V. GROWTH KINETICS AT HIGH SUPERSATURATION LEVELS

A large bulk of the experimental studies on macromolecular crystal growth have been focused at crystal growth relatively close to equilibrium conditions. Consequently, a vast knowledge on protein crystal growth at low-to-moderate supersaturation levels is present. Juxtaposed to this build-up of experimental knowledge is our understanding of protein crystal growth at conditions far away from equilibrium. Many questions concerning the mechanisms and kinetics of crystallization at elevated driving forces remain unanswered. This is, however, contradictory to the fact that the activation barrier for protein nucleation is very large and therefore demands high supersaturation levels to be overcome (for an overview on protein nucleation, we refer to Chapter 4 of this book). Hence, most if not all protein crystals will (in the early postnucleation period) grow at conditions significantly different from near-equilibrium conditions. In this chapter, we therefore discuss crystal growth at high supersaturation levels, where the kinetics of growth alters significantly.

A. Kinetic Roughening

Below the roughening temperature T_R, the overall crystallographic orientation of F-faces is generally maintained. Faceted interfaces are smooth on a molecular scale and growth proceeds by the attachment of new growth units into kink sites along step edges. As mentioned in previous chapters, new step edges are created by screw dislocations intersecting the interface, through the formation of two-dimensional nuclei or through the incorporation of microcrystals [74, 84, 131–134]. If the temperature is equal to or above T_R, the step edge free energy vanishes and the crystal surface becomes rough at a molecular scale and undergoes a thermodynamic phase transition, called *thermal roughening* [53, 71, 135–138]. At a temperature below T_R, the step edge free energy is larger than zero. Consequently, according to the classical nucleation theory [139], a nucleation barrier for the formation of the critical two-dimensional nucleus exists and with it, a critical nucleus size larger than the crystal's individual building blocks. The size of this critical nucleus is inversely proportional to the temperature and the supersaturation. At a constant temperature, an increase in the supersaturation will translate into a decrease in critical nucleus size. As a consequence, in spite of the nonzero step edge free energy, a critical supersaturation σ_R will exist for which the barrier for two-dimensional nucleation essentially vanishes and the size of the critical two-dimensional nucleus is reduced to the order of one growth unit (Fig. 15). Due to a large step density and a very small two-dimensional critical nucleus at elevated supersaturations, the surface becomes rough and offers many favorable sites uniformly distributed across the surface. Consequently, arriving molecules can be incorporated quasi at any site. This transition from a slow, layer-by-layer growth regime to a fast continuous

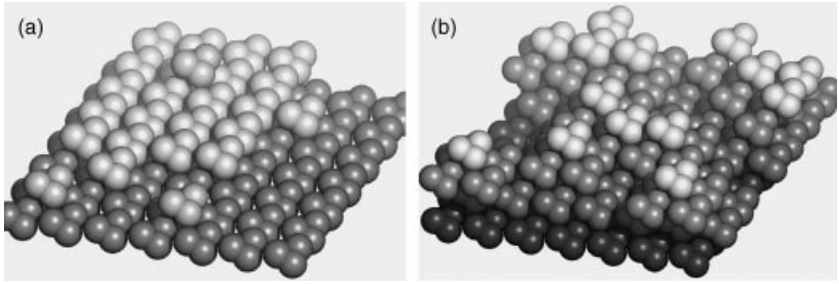


Figure 15. (**a**) Layer-by-layer growth on the (011) face of glucose isomerase; (**b**) continuous growth mode resulting in a rough surface on a molecular scale (growth unit is a tetramer; single sphere is a monomer).

growth regime at high driving forces is called the *kinetic roughening* transition [70, 140–144]. Kinetic roughening for the case of crystallization from solution has been observed for small molecules, that is, SiO_2 [145], Al_2O_3 [146], ZnO and ZnS [147, 148], *n*-paraffins [149, 150], and many others. For the case of protein crystallization, continuous growth has only been hinted at for apoferritin [133] and observed for lysozyme [151–153]. However, these reports are based on average face growth kinetics using low-resolution macroscopic techniques. Here we use *in situ* techniques (AFM, LCM-DIM) to image crystals undergoing the roughening transition in real-time.

Since kinetic roughening is not an actual phase transition, but rather a kinetic phenomenon, it lacks a precise definition [154]. Consensus does exist, however, on the criteria that may be employed to identify it: (i) a critical supersaturation σ_R exists for which the barrier for two-dimensional nucleation essentially vanishes and the theoretical critical radius becomes smaller than half-a-growth unit; (ii) above σ_R face kinetics are expected to follow a linear relationship with the supersaturation; (iii) kinetic roughening coincides with the observed transition from flat to rounded crystal facets. In the following section, we shall shortly address these three criteria for the crystallization of glucose isomerase.

(i) Due to the very small critical size of the two-dimensional nucleus in the kinetic roughening regime, growth units can nucleate in kinks, on step edges and terraces. Growth will therefore proceed by the random addition of solute molecules to the crystal surface. Consequently, the notion of a *step* and a *terrace* is no longer applicable. The onset of this process occurs when the static mean excursion of an isolated step w is larger than the mean terrace width. This is exemplified in Fig. 16a where the interstep distance as a function of supersaturation is plotted. A steady exponential decrease is observed until the interstep distance narrows down to the width of a

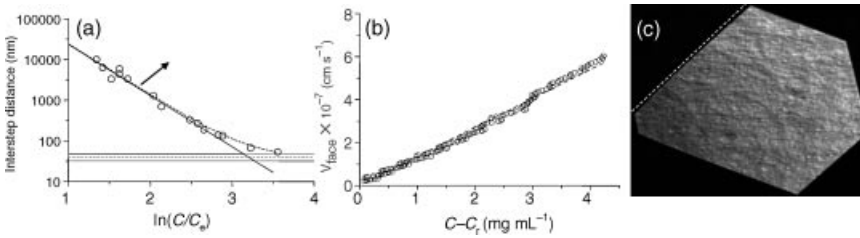


Figure 16. (**a**) Interstep distance as a function of supersaturation (open circles) and guide for eye (dashed line). Best fit was obtained for a first-order exponential decay (full black line). Horizontal dashed line indicates the static mean excursion of an isolated step, full lines show the 95% confidence interval; (**b**) face growth velocities as a function of protein concentration above the roughening concentration C_R; (**c**) continuous growth mode as imaged using LCM-DIM: dashed line reveals very slight crystal rounding.

single step (indicated by horizontal full and dashed lines). AFM stability issues obstruct the determination of the interstep distance at higher supersaturation closer to σ_R. The trend does, however, predict an even greater decline in terrace width, hence corroborating the notion that steps will start to merge and eventually cease to exist. A rough estimate can be made of the number of macromolecules N^* in the critical nucleus at the transition supersaturation σ_R using the following equation [155]:

$$N^* = \frac{\pi \gamma_{\text{step}}^2}{(kT\sigma_R)^2} \tag{11}$$

For $\sigma_R = 5.0 \pm 0.1$ and $\gamma_{\text{step}} = 1.43 \pm 0.01 \times 10^{-20}$ J molecule^{-1} [47], one obtains for $N^* \approx 1.5$. This agrees well with the assumption that the kinetic roughening transition is characterized by a critical nucleus with the size of a few molecules or smaller. Note that although a vanishing critical nucleus size is a clear telltale sign of kinetic roughening, it is not a fundamental prerequisite. Kinetic roughening-like transitions have been observed in both mean-field model and kinetic Monte Carlo simulations where surface roughening was observed at conditions with nonvanishing critical cluster size [156].

(ii) At a temperature and supersaturation below T_R and σ_R, respectively, crystal faces that grow by the multilayer mode (2D nucleation with islands already forming on unfinished layers) are expected to follow a near $\sigma^{5/6}$ dependence, that is, nonlinear [155]. This is indeed observed for glucose isomerase at σ below 5.0. Above the roughening supersaturation, face growth rates R_{face} accelerate due to the absence of a 2D nucleation

activation barrier resulting in

$$R_{\text{face}} = \frac{N_A D a^2}{1.6 M_W \phi^{-1/3}} (C - C_{\text{R}}) \exp(-U_c / kT) \tag{12}$$

with M_{W} is the molecular weight, ϕ is the macromolecular volume fraction, C_{R} is the roughening concentration, and U_{c} is the energy barrier for continuous growth. This model amply predicts the observed face growth rates of glucose isomerase above the roughening transition (Fig. 16b). From the linear kinetics at $\sigma > \sigma_{\text{R}}$, we estimate the energy barrier for continuous growth to be $3.74 \pm 0.01 \times 10^{-20}$ J molecule^{-1} at 20°C which is of the same order as the value obtained for lysozyme ($6.1 \pm 0.4 \times 10^{-20}$ J molecule^{-1}) [152].

(iii) The often polyhedral shape of protein crystals reflects the anisotropy in the surface tension—the orientations present on the habit represent sharp local minima in surface tension. At conditions far from equilibrium, crystal shapes are less determined by thermodynamics and (can) deviate from the equilibrium shape due to kinetic effects. For very fast modes of growth, such as kinetic roughening, isotropic kinetics can alter the anisotropic polyhedral shape. Typically, this is witnessed by a moderate rounding of the crystal faces. This is to a lesser extent also observed for glucose isomerase, although the effect remains quite limited (Fig. 16c).

VI. IMPURITY EFFECTS ON PROTEIN CRYSTAL GROWTH

Proteins are by their very nature difficult to purify to homogeneity and to free from contaminant macromolecules compared to inorganic materials. Even highly purified macromolecules may, for a variety of reasons, be chronically heterogeneous due to modifications, denaturation, the binding of ligands, or a host of other effects. In addition, macromolecules are crystallized from generally complex solutions that include not only the target macromolecule but also buffers, salts, precipitating agents, water, and any number of small effecter molecules [23]. Up to now a large number of studies have demonstrated that the presence of macromolecular impurities during protein crystallization from solution can significantly affect the solubility [157, 158], growth and nucleation kinetics [37, 73, 159–166], morphology [167–171], lattice order [68, 172, 173], and ultimately the diffraction properties of these crystals [88, 174–179].

Generally, impurities are considered to exhibit their effects on the growth process after they adsorb on a crystal surface. Thus, to fully comprehend the mechanisms of impurity effects, one has to observe *in situ* both (1) dynamics of elementary steps and (2) adsorption of impurity molecules on a crystal surface, at a molecular level. Then, the relation between adsorption of impurities and their effects on step

dynamics should be studied thoroughly to gain a better understanding on impurity effects.

To reveal their impurity effects on the growth of a protein crystal, LCM-DIM was used to measure step velocities (Section VI.A) and 2D nucleation rates (Section VI.C) in the presence of various protein impurities to quantify the effects of impurities on the dynamics and kinetics of elementary growth steps. Simultaneously, the adsorption site of fluorescent-labeled impurity molecules (Section VI.B) was observed *in situ* with a single-molecule visualization (SMV) technique [81, 180]. These observations shed light on the direct relation between the preferential adsorption sites of impurities on the crystal surface and their effects on growth kinetics.

A. Deceleration of Step Advancement by Impurities

Two frequently found impurities in commercially available hen egg-white lysozyme preparations (covalently bonded lysozyme dimers and an unidentified protein of 18 kDa molecular weight [181]) and a specially designed microheterogeneous impurity (fluorescent-labeled lysozyme [166]) were intentionally added to highly purified lysozyme solutions to study their influence on the morphology and step velocity of elementary 2D islands on (110) faces of tetragonal lysozyme crystals. Additionally, Seikagaku lysozyme was tested in this study to serve as a commercially available comparative sample.

A first diagnostic feature of the influence of impurities on the growth kinetics is the change of growth step morphology in the presence of impurities in the growth solution. In the case of tetragonal lysozyme crystals 2D islands formed in a 99.99% purity solution (Fig. 17a) are lens shaped with sharp tips. This shape results from an intrinsic anisotropy in the step velocities, being $\langle 001 \rangle$ the slow direction and $\langle 110 \rangle$ the fast direction [32, 182, 183]. The ratio of fast to slow direction is $\sim$6. In Fig. 17a drastic change in 2D island morphology is observed for crystals growing from Seikagaku (98.5% purity [181]) and Sigma lysozyme (94.5% purity [181]). When "purified" impurities (F-lysozyme, dimer, and 18 kDa) are added intentionally to highly purified lysozyme solution, 2D islands exhibit slightly rounded tips. However, note that Seikagaku lysozyme (b), which contains mainly 0.5 wt% dimer and 1.0 wt% 18 kDa as impurities [181] shows significantly stronger effects (the ratio of fast to slow direction is $\sim$3) than the 99.99% purity lysozyme with intentionally added 0.5 wt% dimer and 1.0 wt% 18 kDa (g) (the ratio of fast to slow direction is $\sim$4). This indicates that trace amounts of other impurities included in Seikagaku lysozyme [163, 181] play a significant role.

To assess the effect of impurities on the growth kinetics of tetragonal lysozyme crystals, step velocities were measured in the fast direction $\langle 110 \rangle$ on (110) faces for impurity concentrations ranging from $\langle 0.01 \rangle$ to 2.5% (Fig. 18). In the case of the 99.99% purity solution the step velocity versus supersaturation plot shows

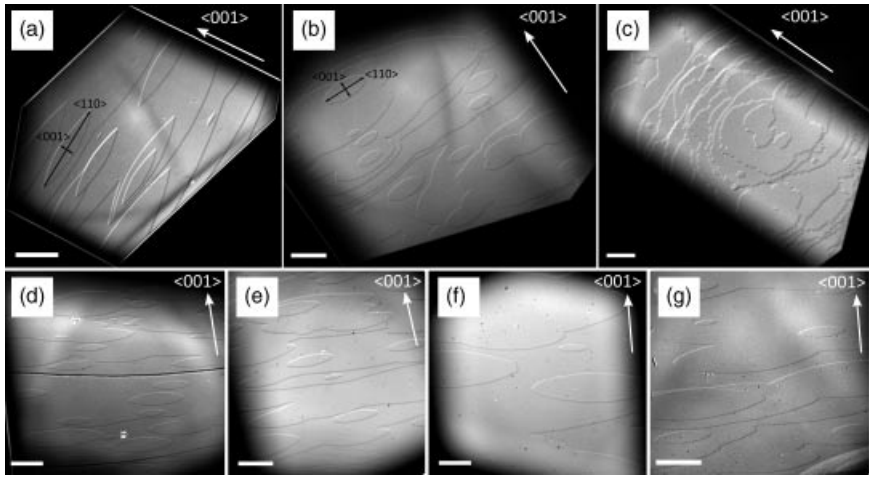


Figure 17. LCM-DIM images of 2D islands on (110) faces of tetragonal lysozyme crystals grown from (**a**) purified lysozyme (99.99% purity), (**b**) Seikagaku lysozyme (98.5% purity), and (**c**) Sigma lysozyme (94.5% purity). In (**d**–**g**), all crystals were grown from 99.99% purity lysozyme solutions with intentionally added impurities: (**d**) 0.1 wt% F-lysozyme, (**e**) 0.5 wt% dimer, (**f**) 1.0 wt% 18 kDa, and (**g**) 0.5 wt% dimer and 1.0 wt% 18 kDa. Scale bars represent 20 μm.

a slightly concave shape, in particular under a low supersaturation range. Since kink density of a 2D island in the fast direction is high, nucleation of kinks (one-dimensional nucleation) cannot be the rate-determining step of the growth process. Hence, the concave shape of the step velocity versus supersaturation plot is most likely due to impurities present in the purified solution (>99.99%). A similar impurity effect was observed by Vekilov and Rosenberger for a 99.99% pure lysozyme solution in forced flow experiments [184]. Figure 18 demonstrates that all three protein impurities decreased the step velocity, as well as the case of Seikagaku lysozyme. With increasing impurity concentration, the suppression of the step velocity increases, and the shape of the plots becomes more concave. In particular, the suppression of the step velocity by 18 kDa and impurities included in Seikagaku lysozyme are significant. To simulate Seikagaku lysozyme, 0.5 wt% dimer and 1.0 wt% 18 kDa were used together with 99.99% purity lysozyme. However, the suppression of this solution (open triangles) was not as strong as that observed for Seikagaku lysozyme, which implies that trace amounts of other impurity(ies) present in Seikagaku lysozyme contribute to the larger suppression of the step velocity.

To quantitatively evaluate the effects of these protein impurities, the ratios R of the step velocity in the quasi *pure* to the impure case was calculated at the same supersaturation. R thus quantifies the suppressive capacity of these impurities

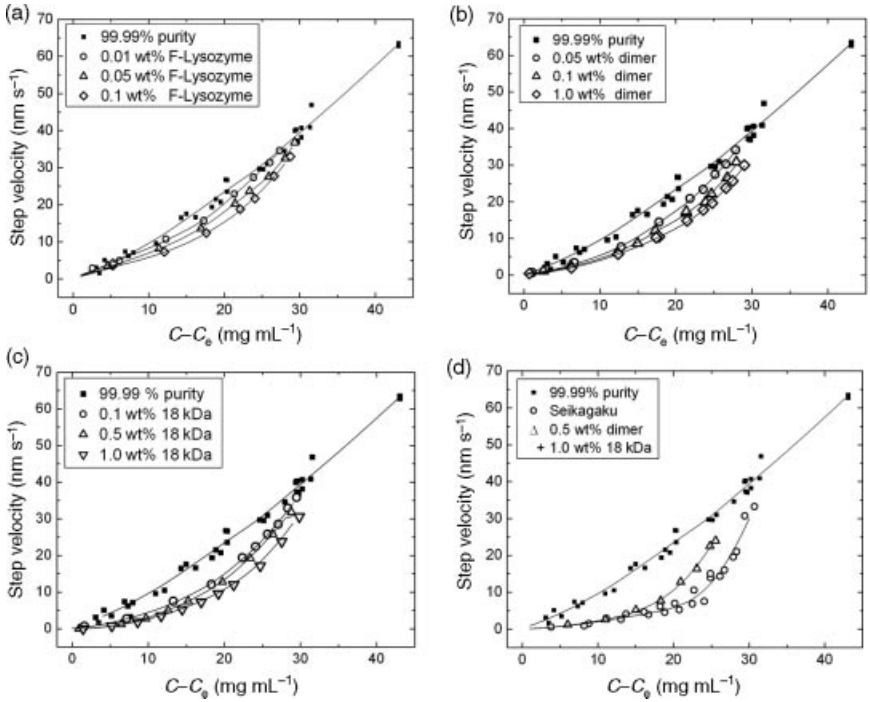


Figure 18. Step velocities in ⟨110⟩ direction on (110) faces of tetragonal lysozyme crystals at various supersaturations and protein impurity concentrations. All crystals were grown from 99.99% purity lysozyme solutions of 40 mg mL^{-1}. Protein impurities added intentionally: (**a**) 0.01–0.1 wt% F-lysozyme, (**b**) 0.05–1.0 wt% dimer, (**b**) 0.1–1.0 wt% 18 kDa, and (**c**) 0.5 wt% dimer and 1.0 wt% 18 kDa. In (d), also the result of Seikagaku lysozyme (98.5% purity) solutions of 40 mg mL^{-1} was plotted. The data points were fitted using a local polynomial regression [185].

on step kinetics. As shown in Fig. 19, R is supersaturation, impurity type, and concentration dependent. For instance, the R values of 18 kDa are significantly larger than those of F-lysozyme and the dimer impurity. For the same impurity, however, the shape of the supersaturation dependence plots does not change with impurity concentration. We discern three qualitatively different dependencies: (i) R displays a plateau at low supersaturation and at higher supersaturation decreases with increasing supersaturation (F-lysozyme), (ii) R decreases monotonically as a function of supersaturation (dimer and 18 kDa), and (iii) a mixed regime characterized by a steady decrease with a local maximum of R (Seikagaku). Note that the alleged Seikagaku-mimic, that is, 0.5% dimer and 1.0% 18 kDa was not able to reproduce the Seikagaku lysozyme plot, indicative of an overseen presence of (trace amounts of) additional impurities.

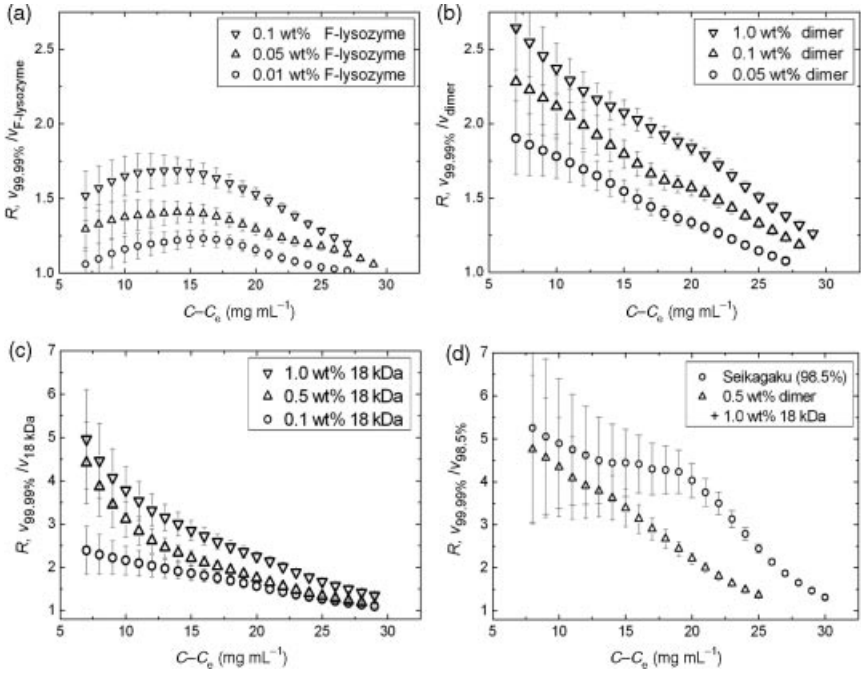


Figure 19. Supersaturation dependence of R for various impurity types. Errors were evaluated from the standard error (90% statistical significance) and the propagation of errors. (**a**) 0.01–0.1 wt% F-lysozyme, (**b**) 0.05–1.0 wt% dimer, (**c**) 0.1–1.0 wt% 18 kDa, and (**d**) 0.5 wt% dimer with 1.0 wt% 18 kDa, Seikagaku lysozyme.

B. Impurity Adsorption Sites

To rationalize these dependencies, it is imperative to obtain information on the preferential sites of adsorption of the impurities discussed above. However, accurate identification of those adsorption sites requires molecular resolution imaging combined with the ability to discern between the abundant growth species and the rare but pernicious impurities. Ideal for this purpose is the technique, total internal reflection fluorescence (TIRF) microscopy of the thin-solution-layer type [81], which allows SMV. Combined with LCM-DIM, one can pinpoint the location of fluorescent species (here, impurities) and image elementary steps on identical field-of-views. Selected for such a study are two of the model impurities exhibiting the different supersaturation dependencies, that is, F-lysozyme (class (i)) [166] and fluorescent-labeled lysozyme dimers (F-dimer, class (ii)) [180]. The latter displays the same impurity characteristics as the nonfluorescently labeled dimers and therefore makes for a more useful and valid analog [186]. The main results are

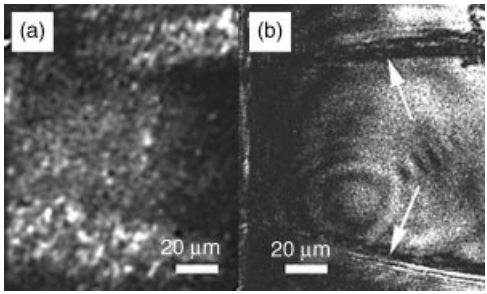


Figure 20. F-lysozyme adsorbs preferentially on steps on the (110) surface of a tetragonal lysozyme crystal. (**a**) Single-molecule TIRF image reveals that the majority of F-lysozyme molecules are located in areas of high step density. (**b**) This becomes readily apparent when compared to the same field-of-view imaged by LCM-DIM, from which the position of bunched steps can be inferred. We note that these step bunches correspond to regions of high fluorescence and thus F-lysozyme surface concentration.

summarized in Figs. 20 and 21 for F-lysozyme and F-dimer, respectively. Figures 20a and 21a show typical single-molecule TIRF images of a lysozyme crystal surface exposed to one of these impurities. One bright dot corresponds to one F-lysozyme/F-dimer molecule adsorbed on the crystal surface (higher intensity dots correspond to overlapping of fluorescence spots). Figures 20b and 21b represent the same field of view observed by LCM-DIM showing elementary and bunched (indicated by arrows). The comparison between images 20/21a and 20/21b clearly demonstrates the preferential adsorption of F-lysozyme on steps and the at random adsorption of F-dimers on the terraces in between the steps.

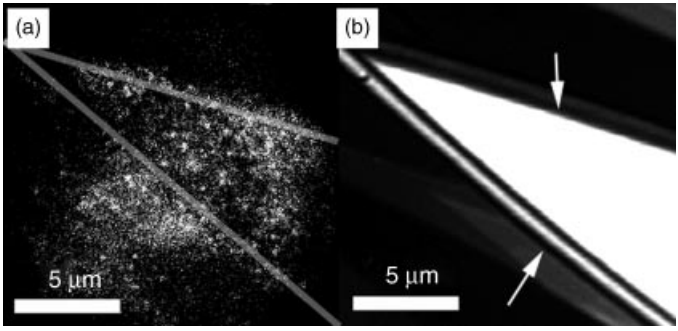


Figure 21. F-dimers adsorb preferentially on terraces on the (110) surface of a tetragonal lysozyme crystal. (**a**) Single-molecule TIRF image reveals a quasiuniform distribution of F-dimers on the surface. (**b**) Comparison to the same field-of-view imaged by LCM-DIM shows that fluorescence spots do no correlate to the location of (macro)steps.

What could cause such a difference in adsorption preference? To answer this, one needs to take into account that a lysozyme dimer is composed of two covalently bonded lysozyme molecules [181] with a bonding orientation that differs from the symmetry between the two neighboring lysozyme molecules within the crystal lattice [187]. Hence, it will be less favorable for a dimer to fit into a kink site because it will suffer large steric hindrance and unfavorable bonding with its first and possibly second neighbors. In contrast, steric restraints are relaxed on the open areas of a terrace resulting in a higher probability of occurrence of dimers on terraces as opposed to kinks. Monomeric F-lysozyme species do not suffer from these incompatibilities to the same extent. The minute change in mass and volume (3% mass increase compared to native lysozyme) do not render these labeled molecules incompatible with the lattice. F-lysozyme will therefore be enriched in sites of minimal energy, that is, at step edges or within kinks. The results allow us to conclude that protein impurities whose intermolecular bonding to the crystal surface is close to that of a solute molecule will preferentially adsorb on steps and kinks, whereas impurity molecules with substantially different interaction potentials will preferentially adsorb on sites less sensitive to a perfect fit, that is, terraces.

C. Adsorption Sites, Supersaturation Dependencies, and Relevant Timescales

From the difference in the adsorption preference (step or terrace) we can tentatively formulate a model that accounts for the observed R versus $C-C_e$ dependencies. Impurity models can essentially be divided into two main groups: (i) impurities that adsorb at kink sites leading to kink blocking, proposed by Bliznakov [188], and presented more extensively by Chernov [40]; (ii) impurities adsorbing on terraces (and/or steps) leading to pinning of steps first described in the pioneering work of Cabrera and Vermileya [120].

First, we address impurities that target kinks. At low-to-moderate supersaturation levels, steps advance by the incorporation of solute molecules into kinks. Consequently, step advancement is directly proportional to the density of kinks within that step. Impurities that attach to the kink and preclude (or slow down) any further solute attachment (kink blocking) effectively lower the density of kinks that can contribute to step advancement, thus retarding the step [189]. Impurity attachment to the kink is a kinetic event marked by a characteristic timescale referred to as τ_{imp}. Similarly, a supersaturation-dependent kink exposure time $\tau_{kink}(\sigma)$ exists that can be defined as the average time between kink creation and solute incorporation into the kink. At low supersaturation, $\tau_{kink} \gg \tau_{imp}$, thus enabling the impurity to reach equilibrium density N_{1D} within the steps (Fig. 22a), leading to maximal impurity effect R for a given impurity concentration (Fig. 22c). The step kinetic coefficient β_{step} reaches its minimal value leading to a reduced slope in the

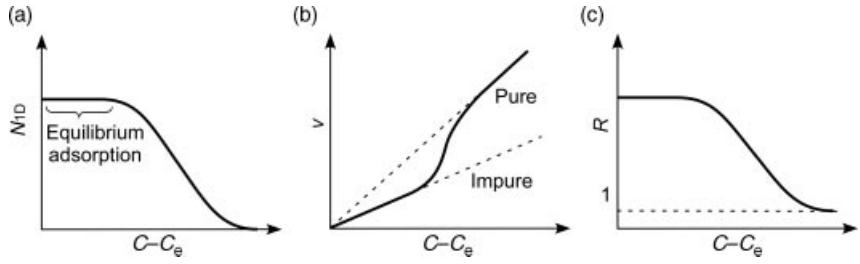


Figure 22. Schematic illustrations of impurity effects of F-lysozyme. Supersaturation dependencies of number density N_{1D} (m^{-1}) of adsorbed F-lysozyme on a step (**a**), step velocity v (**b**), and the impurity effect R (**c**).

v_{step} versus C–C_e plots (Fig. 22b, impure). If the supersaturation is increased, a crossover regime is reached where the characteristic timescales of impurity binding and kink exposure start to overlap, that is, $\tau_{kink} \sim \tau_{imp}$. Impurity densities within the step do not reach equilibrium value, less kinks are blocked and the impurity effect diminishes. Until finally, a third regime is reached for higher supersaturation where $\tau_{kink} \ll \tau_{imp}$, here impurity binding is abolished by the incompatibility of timescales. Kink densities reach their value as obtained in the pure case and the impurity effect on step kinetics becomes unsignificant (Fig. 22b, pure; Fig. 22c, $R = 1$). This model amply explains the R versus $C-C_e$ plots observed for F-lysozyme (a "kink-impurity") and can be visualized as follows: maximal step deceleration at low supersaturation followed by a steady increase in step velocity to nominal values toward higher supersaturation. Note that the width of the plateau in Fig. 22c depends on the initial difference between τ_{kink} and τ_{imp}. However, limiting case may exist where *slow* impurities never reach equilibrium densities within the step. For these cases, no plateau in R versus $C-C_e$ is expected. We also stress that the supersaturation dependencies presented here are for cases where kink formation is not the limiting step. Steps with very low kink densities, show considerably different supersaturation dependencies of the impurity effect [190].

Next we discuss the case of the lysozyme dimer. Similarly to the adsorption preference of F-dimer on the (110) face (discussed above), AFM observations have revealed that terraces are also the major adsorption sites of the nonfluorescently labeled dimers on the (101) face [164]. To describe the impurity effects of dimer on step advancement, we follow the step pinning model [120], which attributes step retardation to the Gibbs–Thomson effect for curved steps. Under low supersaturation, layer generation is slow and step velocities are small, yielding high terrace exposure times $\tau_{terrace}$. The dimer will show equilibrium adsorption and reach its maximal surface density N_{2D} ($\tau_{terrace} \gg \tau_{imp}$). For high dimer surface densities, the average distance d between the adsorbed impurities will be small. When d is smaller than the diameter $2\rho_{crit}$ of a critical 2D nucleus, steps stop advancing

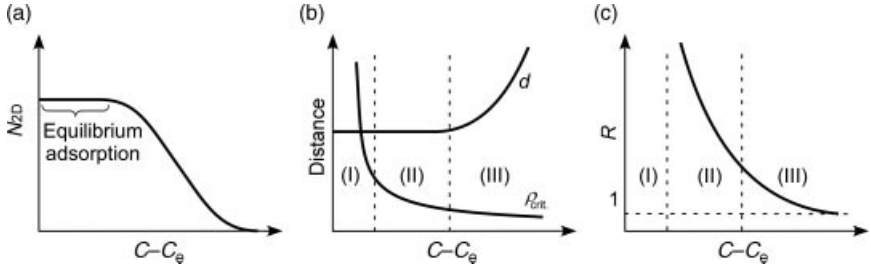


Figure 23. Schematic illustrations of impurity effects of dimer. Supersaturation dependencies of number density N_{2D} (m^{-2}) of adsorbed dimer on terraces (**a**), distance d between adjacent dimer molecules adsorbed on a terrace and radius ρ_{crit} of a critical 2D nucleus (**b**) and the impurity effect R (**c**).

(supersaturation range I in Fig. 23b) and R cannot be defined. However, in the supersaturation range II (Fig. 23b), where equilibrium adsorption of the dimer still exists ($\tau_{terrace} \gg \tau_{imp}$), with increasing supersaturation a decrease in $\rho_{crit.}$ results in a decrease of the impurity effect (i.e., R). In the supersaturation range III, terrace exposure times decrease and $\tau_{terrace}$ does become comparable to τ_{imp}. As a result, both d and ρ_{crit} decrease thus diminishing the impurity effect R. The supersaturation range observed in Fig. 23b corresponds to ranges II and III in Fig. 23c. We stress here again that depending on the initial values of $\tau_{terrace}$ and τ_{imp} a significant change of the width of the regimes may occur. For instance, if $\tau_{terrace}$ is equal to or smaller than τ_{imp} even for very low supersaturation then regimes I and II will be significantly narrower. Similar terrace exposure time-dependent impurity effects have also been observed for, for example, calcium oxalate monohydrate [191].

D. Impurity Effect on 2D Nucleation

Just as in the case of step advancement, 2D nucleation rates were measured for lysozyme crystal growth under a wide supersaturation range in the presence of three different impurities: F-lysozyme, covalently bonded dimer of lysozyme (dimer) and the 18 kDa polypeptide. These impure proteins were intentionally added to highly purified lysozyme solutions (99.99%). Similar to the step kinetics study, 2D nucleation rates are also measured from Seikagaku lysozyme solutions to serve as a commercially available impurity model system.

Figure 24 shows changes in $\ln J$ as a function of $1/[T^2 \ln(C/C_e)]$. A clear linear relationship is found for all the data points under a higher supersaturation range $\sigma > 0.80$ (solid line). This result demonstrates clearly that impurities do not affect 2D nucleation rates in a high supersaturation range (i.e., homogeneous nucleation). However, the plots of lysozyme with impurities exhibit a kink at the middle supersaturation $\sigma \approx 0.80$, leading to a much smaller slope under a lower supersaturation range $\sigma < 0.80$ (dash-dotted line). This smaller slope indicates the occurrence of heterogeneous 2D nucleation in this lower supersaturation range. Heterogeneous

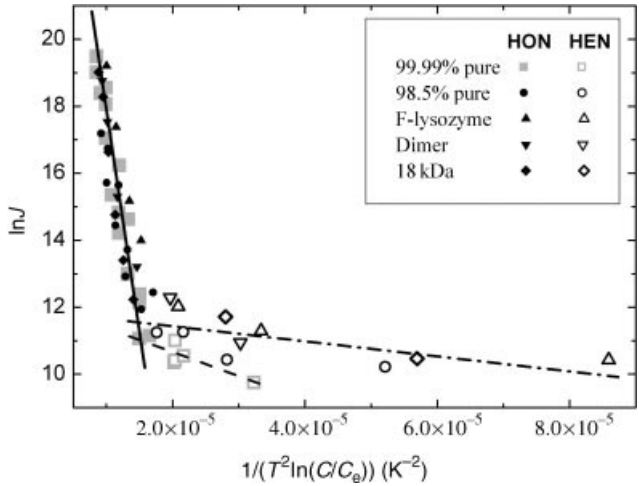


Figure 24. Changes in 2D nucleation rates as a function of $1/[T^2\ln(C/C_e)]$, measured on (110) faces of tetragonal lysozyme crystals. The black line represent homogeneous 2D nucleation (HON) under a higher supersaturation range, the dashed line represents heterogeneous 2D nucleation (HEN) in the purified lysozyme solution under a lower supersaturation range, and the dash-dotted line represents HEN in solutions containing impurities under a lower supersaturation range.

2D nucleation under a low supersaturation range in protein crystallization has also been reported for thaumatin [55, 56], catalase [56], and glucose isomerase [38].

Liu and Maiwa [192] gave a theoretical explanation and experimental examples for the heterogeneous 2D nucleation on inorganic crystals. This model assumes that any kind of foreign particles (e.g., impurity molecules), can potentially serve as nucleation centers for 2D nucleation, and they promote nucleation by lowering the 2D nucleation barriers. This model is confirmed when looking at the edges free energies of the (110) face obtained from measuring the 2D nucleation rates as a function of supersaturation and using Eq. (8). Table II shows that the edge

TABLE II
Edge Free Energies κ ($J\,m^{-1}$) and Surface Free Energies α ($mJ\,m^{-2}$) Obtained by the Linear Curve Fittings of the Experimental Data Shown in Fig. 24 Using Eq. (8)

	κ ($J\,m^{-1}$)	α ($mJ\,m^{-2}$)
HON	$(3.5 \pm 0.2) \times 10^{-12}$	0.62 ± 0.03
HEN 99.99% pure	$(0.9 \pm 0.2) \times 10^{-12}$	0.16 ± 0.03
HEN impurities	$(0.5 \pm 0.1) \times 10^{-12}$	0.09 ± 0.02

α was calculated as $\alpha = \kappa/h$, where h is a step height; HON and HEN denote homogeneous and heterogeneous 2D nucleation, respectively.

free energy in the low supersaturation range is significantly lower than that at high supersaturations. Hence, the cost of creating a 2D nucleus is lower at low supersaturations than at higher supersaturations as predicted by the model of Liu. The ledge free energy of the heterogeneous nucleation for the 99.99% pure solution is larger than that for the solutions intentionally containing impure proteins. This would result from the difference in the impurity concentrations.

VII. CONCLUDING REMARKS

The past two decades, experimental protein crystal growth studies have been focused on growth phenomena operating on either a microscopic or macroscopic scale, using atomic force microscopy and phase shifting Michelson interferometry, respectively. The former has allowed the investigation of nanoscale processes (defect formation, nucleus characterization, growth mechanisms and growth kinetics, etc.) with molecular-level resolution on a (highly) local scale, while the latter has been successful at elucidating crystal-scale phenomena (normal growth and surface morphology, coupling of incorporation kinetics and mass transport, step bunching and growth instabilities, etc.). Processes that operate at the intermediate level and thus require the combination of a crystal-scale observation area and the ability to discern elementary steps on the surface have remained largely unexplored. This can be attributed due to limited accessibility to such regimes by the techniques present at the time. *In situ* analysis of these mesoscopic processes is, however, vital in linking elementary incorporation processes at the kink-level to macroscopic crystal properties such as crystal face kinetics, diffraction characteristics, crystal morphology, and so on. This mesoscopic level has become fully accessible to the experimentalist with the advent of laser confocal differential interference microscopy. Additionally, the combination of the relatively new LCM-DIM with AFM and PSMI has proven to be quite powerful as it allows the user to investigate protein crystal growth at all lengthscales, the micro-, meso-, and macroscopic level. To summarize, from *in situ* mesoscopic scale observations of protein crystal growth we have learned and/or found direct experimental evidence that are as follows:

- proteins can also grow by the 2D nucleation mechanism even at low supersaturation due to the lack of active spirals on the crystal face;
- regimes exist where the destabilizing action of the Berg effect greatly exceeds crystal face stability mechanisms leading to a profound change of protein crystal morphology;
- at high supersaturation levels, the activation barrier for 2D nucleation becomes comparable to kT leading to a vanishing critical nucleus size marking the onset of the kinetic roughening regime;

- the adsorption site of impurities will dictate the impurity effect on step advancement as a function of supersaturation;
- even very small impurity concentration can have very fast effects on the surface morphology;
- impurities induce a significant increase in 2D nucleation rates in a low supersaturation range.

In this overview, we have demonstrated that LCM-DIM is a very useful technique for *in situ* observations of protein crystal growth at a mesoscopic level connecting atomic and macroscopic observations. LCM-DIM allows for noninvasive and *in situ* observation of step dynamics and surface processes on crystals growing from solution. These characteristics make this technique very useful for rapid, and precise, determination of protein crystal solubility [193] and the search of ideal growth conditions for obtaining very large crystals necessary for neutron diffraction. Some very promising applications of this technique are the study of 2D protein growth and determination of growth dynamics of membrane protein crystals. Apart from the benefits of LCM-DIM reported in this overview, studying *in situ* protein crystal growth processes taking place under experimental conditions which are not accessible by AFM or PSI becomes possible, such as crystal growth in gelled solutions [50] or under high pressure [194]. With LCM-DIM, one can also collect information from inside the crystal volume making it possible to visualize *in situ* the strain fields around the dislocations with sufficient contrast and higher resolution than possible with X-ray topography [36].

Next to the observation of step dynamics on protein surfaces, this technique can also be used for small molecule crystals, such as monoclinic gypsum crystals (step height $\approx$ 0.7–0.8 nm) [195], ultra-flat Au(111) surfaces (step height $\sim$ 0.25 nm) [196], and even ice crystals (step height $\sim$ 0.37 nm) [197]. Another promising application is the combination of LCM-DIM with other observation methods, such as AFM [198], yielding simultaneous observation of steps on a crystal scale and height determination of the surface features. A second, very powerful, combined observation method is LCM-DIM linked to a single molecule observation setup which opens the door to study diffusion processes of single molecules on the crystal surface [81] and to study the attachment–detachment dynamics of single molecules *in situ* [199] during crystal growth.

Acknowledgments

We thank T. Nakada for making some unpublished results available. We are greatly indebted to P. G. Vekilov, J. A. Gavira, F. Otalora, J. M. Garcia-Ruiz, K. Decanniere, G. Nicolis, K. Tsukamoto, and G. Sazaki for stimulating discussions on many aspects of this work. The financial support by the European Space Agency under contract number ESA AO-2004-070 is also greatly appreciated. Alexander Van Driessche is grateful for the support by the Consolider-Ingenio 2010 project "Factoría Española de cristalización."

References

1. N. Deshpande, K. J. Addess, W. F. Bluhm, J. C. Merino-Ott, W. Townsend-Merino, Q. Zhang, C. Knezevich, L. Xie, L. Chen, Z. Feng, R. K. Green, J. L. Flippen-Anderson, J. Westbrook, H. M. Berman, and P. E. Bourne, *Nucleic Acids Res*,. **33**, D233–D237 (2005).
2. S. K. Burley, S. C. Almo, J. B. Bonanno, M. Capel, M. R. Chance, T. Gaasterland, D. Lin, A. Sali, F. W. Studier, and S. Swaminathan, *Nat. Genet.*, **23**, 151 (1999).
3. J. P. K. Doye and W. C. K. Poon, *Curr. Opin. Colloid Interface Sci.*, **11**, 40 (2006).
4. J. P. Doye, A. A. Louis, and M. Vendruscolo, *Phys. Biol.*, **1**, 9 (2004).
5. J. A. Jedziniak, J. H. Kinoshita, E. M. Yates, L. O. Hocker, and G. B. Benedek, *Invest. Ophthalmol.*, **11**, 905 (1972).
6. R. J. Siezen, M. R. Fisch, C. Slingsby, and G. B. Benedek, *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 1701 (1985).
7. J. A. Thomson, P. Schurtenberger, G. M. Thurston, and G. B. Benedek, *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 7079 (1987).
8. P. G. Vekilov, A. R. Feeling-Taylor, D. N. Petsev, O. Galkin, R. L. Nagel, and R. E. Hirsch, *Biophys. J.*, **83**, 1147 (2002).
9. O. Galkin, K. Chen, R. L. Nagel, R. E. Hirsch, and P. G. Vekilov, *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 8479 (2002).
10. W. A. Eaton and J. Hofrichter, *Advances in Protein Chemistry*, Academic Press, San Diego, 1990.
11. P. M. Colman, E. Suzuki, and A. van Donkelaar, *Eur. J. Biochem.*, **103**, 585 (1980).
12. M. A. Giembycz and M. A. Lindsay, *Pharmacol. Rev.*, **51**, 213 (1999).
13. M. Veenhuis, J. A. K. W. Kiel, and I. J. van der Klei, *Microsc. Res. Tech.*, **61**, 139 (2003).
14. D. Goulson, *Oecologica*, **109**, 219 (1997).
15. G. Jedd and N.-H. Chua, *Nat. Cell. Biol.*, **2**, 226 (2000).
16. G. Dodson and D. Steiner, *Curr. Opin. Struct. Biol.*, **8**, 189 (1998).
17. S. Matsuda, T. Senda, S. Itoh, G. Kawano, H. Mizuno, and Y. Mitsui, *J. Biol. Chem.*, **264**, 13381 (1989).
18. P. G. Vekilov and A. A. Chernov, *Solid State Phys*,. **57**, 2 (2002).
19. C. N. Nanev, *Cryst. Growth Des.*, **7**, 1533 (2007).
20. P. G. Vekilov, *Cryst. Growth Des.*, **4**, 671 (2004).
21. J. F. Lutsko and G. Nicolis, *Phys. Rev. Lett.*, **96**, 461021 (2006).
22. A. McPherson, A. J. Malkin, and Y. G. Kuznetsov, *Structure*,. **3**, 759 (1995).
23. A. McPherson, *Crystallization of Biological Macromolecules*, Cold Spring Harbor Laboratory Press, New York, 1999.
24. P. G. Vekilov and J. I. D. Alexander, *Chem. Rev.*, **100**, 2061 (2000).
25. A. A. Chernov, *J. Mater. Sci. Mater.*, **12**, 437 (2001).
26. A. McPherson, Y. G. Kuznetsov, A. Malkin, and M. Plomp, *J. Struct. Biol.*, **142**, 32 (2003).
27. A. A. Chernov, *J. Struct. Biol.*, **142**, 3 (2003).
28. A. A. Chernov, *J. Cryst. Growth*, **264**, 499 (2004).
29. P. G. Vekilov, *Cryst. Growth Des.*, **7**, 2796 (2007).
30. P. G. Vekilov, *Cryst. Growth Des.*, **7**, 2239 (2007).
31. M. L. Pusey and R. Naumann, *J. Cryst. Growth*, **76**, 593 (1986).
32. S. D. Durbin and W. E. Carlson, *J. Cryst. Growth*, **122**, 71 (1992).

33. Y. G. Kuznetsov, A. J. Malkin, A. Greenwood, and A. McPherson, *J. Struct. Biol.*, **114**, 184 (1995).
34. O. Gliko, N. A. Booth, E. Rosenbach, and P. G. Vekilov, *Cryst. Growth Des.*, **2**, 381 (2002).
35. G. Sazaki, T. Matsui, K. Tsukamoto, N. Usami, T. Ujihara, K. Fujiwara, and K. Nakajima, *J. Cryst. Growth*, **262**, 536 (2004).
36. G. Sazaki, K. Tsukamoto, S. Yai, M. Okada, and K. Nakajima, *Cryst. Growth Des.*, **5**, 1729 (2005).
37. A. E. S. Van Driessche, G. Sazaki, F. Otalora, F. M. Gonzalez-Rico, P. Dold, K. Tsukamoto, and K. Nakajima, *Cryst. Growth Des.*, **7**, 1980 (2007).
38. M. Sleutel, R. Willaert, L. Wyns, and D. Maes, *Cryst. Growth Des.*, **9**, 497 (2009).
39. K. Tsukamoto and P. Dold, *Perspectives on Inorganic, Organic, and Biological Crystal Growth: From Fundamentals to Applications*, American Institute of Physics, New York, 2007.
40. A. A. Chernov, *Modern Crystallography III. Crystal Growth*, Springer-Verlag, Berlin, 1984.
41. P. Hartman and W. G. Perdok, *Acta Cryst*,. **8**, 49 (1955).
42. P. Hartman, *Crystal Growth: An Introduction*, North-Holland, Amsterdam, 1973.
43. I. V. Markov, *Crystal Growth for Beginners: Fundamentals of Nucleation, Crystal Growth and Epitaxy*, World Scientific Publishing, Singapore, 2003.
44. N. Kitamura, M. G. Lagally, and M. B. Webb, *Phys. Rev. Lett.*, **71**, 2082 (1993).
45. T. A. Land, J. J. De Yoreo, and J. D. Lee, *Surf. Sci.*, **384**, 136 (1997).
46. J. P. Astier, D. Bokern, L. Lapena, and S. Veesler, *J. Cryst. Growth*, **226**, 294 (2001).
47. M. Sleutel, D. Maes, L. Wyns, and R. Willaert, *Cryst. Growth Des.*, **8**, 4409 (2008).
48. T. A. Land, A. J. Malkin, Y. Kutznesov, A. McPherson, and J. J. De Yoreo, *J. Cryst. Growth*, **166**, 893 (1996).
49. O. Gliko, N. Neumaier, W. Pan, I. Haase, M. Fischer, A. Bacher, S. Weinkauf, and P. G. Vekilov, *J. Am. Chem. Soc.*, **127**, 3433 (2005).
50. A. E. S. Van Driessche, F. Otalora, G. Sazaki, M. Sleutel, K. Tsukamoto, and J. A. Gavira, *Cryst. Growth Des.*, **8**, 4316 (2008).
51. D. Gasperino, A. Yeckel, B. K. Olmsted, M. D. Ward, and J. J. Derby, *Langmuir*, **22**, 6578 (2006).
52. K. Chen and P. G. Vekilov, *Phys. Rev. E*, **66**, 021606 (2002).
53. A. J. Malkin, A. A. Chernov, and I. V. Alexeev, *J. Cryst. Growth*, **97**, 765 (1989).
54. K. Maiwa, K. Tsukamoto, and I. Sunagawa, *Proceedings of the Fourth Topical Meeting on Crystal Growth Mechanism*, Tokyo, 1991, pp. 67–70.
55. A. J. Malkin, Y. G. Kuznetsov, W. Glantz, and A. McPherson, *J. Phys. Chem.*, **100**, 11736 (1996).
56. A. J. Malkin, Y. G. Kuznetsov, and A. McPherson, *J. Cryst. Growth*, **196**, 471 (1999).
57. A. J. Malkin and R. E. Thorne, *Methods*, **34**, 273 (2004).
58. G. K. Christopher, A. G. Phipps, and R. J. Gray, *J. Cryst. Growth*, **191**, 820 (1998).
59. P. Atkins, *Physical Chemistry*, 6th ed., Freeman, New York, 1998.
60. Y. B. Yu, *J. Pharm. Sci.*, **90**, 2099 (2001).
61. Y. Suzuki, G. Sazaki, K. Visuri, K. Tamura, K. Nakajima, and S. Yanagiya, *Cryst. Growth Des.*, **2**, 321 (2002).
62. G. Sazaki, K. Kurihara, T. Nakada, S. Miyashita, and H. Komatsu, *J. Cryst. Growth*, **169**, 355 (1996).
63. R. J. Gray, W. B. Hou, A. B. Kudryavtsev, and L. J. DeLucas, *J. Cryst. Growth*, **232**, 10 (2001).
64. T. Ihle, C. Misbah, and O. Pierre-Louis, *Phys. Rev. B*, **58**, 2289 (1998).

65. M. Giesen-Seibert, R. Jentjens, M. Poensgen, and H. Ibach, *Phys. Rev. Lett.*, **71**, 3521 (1993).
66. N. C. Bartelt, J. L. Goldberg, T. L. Einstein, and E. D. Williams, *Surf. Sci.*, **273**, 252 (1992).
67. D. Williams and M. C. Bartelt, *Science*, **251**, 393 (1991).
68. S. T. Yau, B. R. Thomas, and P. G. Vekilov, *Phys. Rev. Lett.*, **85**, 353 (2000).
69. D. N. Petsev, K. Chen, O. Gliko, and P. G. Vekilov, *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 792 (2003).
70. H. E. Buckley, *Crystal Growth*, Wiley, New York, 1951.
71. W. K. Burton, N. Cabrera, and F. C. Frank, *Philos. Trans. R. Soc. Lond.*, **243**, 299 (1951).
72. P. G. Vekilov, *Methods in Molecular Biology,* Vol. 300: *Protein Nanotechnology, Protocols, Instrumentation and Applications*, Humana Press Inc., Totowa, NJ, 2005.
73. P. G. Vekilov and F. Rosenberger, *J. Cryst. Growth*, **158**, 540 (1996).
74. I. Reviakine, *J. Am. Chem. Soc.*, **125**, 11684 (2003).
75. P. G. Vekilov, L. A. Monaco, and F. Rosenberger, *J. Cryst. Growth*, **156**, 267 (1995).
76. P. G. Vekilov, B. R. Thomas, and F. Rosenberger, *J. Phys. Chem. B*, **102**, 5208 (1998).
77. A. A. Chernov, L. N. Rashkovich, I. V. Yaminski, and N. V. Gvozdev, *J. Phys. Condens. Matter*, **11**, 9969 (1999).
78. A. A. Chernov and H. Komatsu, *Science and Technology of Crystal Growth*, Kluwer Academic, Dordrecht, 1995.
79. A. A. Chernov and H. Komatsu, *Science and Technology of Crystal Growth*, Kluwer Academic, 1995.
80. M. Sleutel, C. Vanhee, C. Van de Weerdt, K. Decanniere, D. Maes, L. Wyns, and R. Willaert, *Cryst. Growth Des.*, **8**, 1173 (2008).
81. G. Sazaki, M. Okada, T. Matsui, T. Watanabe, H. Higuchi, K. Tsukamoto, and K. Nakajima, *Cryst. Growth Des.*, **8**, 2024 (2008).
82. A. A. Chernov, *Acta Crystallogr. A* **54**, 859 (1998).
83. R. A. Garza-López, P. Bouchard, G. Nicolis, M. Sleutel, J. Brzezinski, and J. Kozak, *J. Chem. Phys.*, **128**, 114701 (2008).
84. A. McPherson, *Crystallization of Biological Macromolecules*, Cold Spring Harbor Laboratory Press, New York, 1999.
85. B. A. Nerad and P. J. Shlichta, *J. Cryst. Growth*, **75**, 591 (1986).
86. L. A. Monaco and F. Rosenberger, *J. Cryst. Growth*, **129**, 465 (1993).
87. P. G. Vekilov, J. I. Alexander, and F. Rosenberger, *Phys. Rev. E*, **54**, 6650 (1996).
88. D. C. Carter, K. Lim, J. X. Ho, B. S. Wright, P. D. Twigg, T. Y. Miller, J. Chapmann, K. Keeling, J. Ruble, P. G. Vekilov, B. R. Thomas, F. Rosenberger, and A. A. Chernov, *J. Cryst. Growth*, **196**, 623 (1999).
89. B. R. Thomas, A. A. Chernov, P. G. Vekilov, and D. C. Carter, *J. Cryst. Growth*, **211**, 149 (2000).
90. A. A. Chernov, J. M. Garcia-Ruiz, and B. R. Thomas, *J. Cryst. Growth*, **232**, 184 (2001).
91. E. H. Snell, *Acta Crystallogr. D Biol. Crystallogr.*, **52**, 529 (1996).
92. F. Otalora, M. L. Novella, J. A. Gavira, B. R. Thomas, and J. A. Gavira, *Acta Crystallogr. D*, **57**, 412 (2001).
93. F. Otalora, J. M. Garcia-Ruiz, L. Carotenuto, D. Castagnolo, M. L. Novella, and A. A. Chernov, *Acta Crystallogr. D*, **58**, 1681 (2002).
94. J. M. Garcia-Ruiz, M. L. Novella, and F. Otalora, *J. Cryst. Growth*, **196**, 703 (1999).
95. J. M. Garcia-Ruiz, M. L. Novella, R. Moreno, and J. A. Gavira, *J. Cryst. Growth*, **232**, 165 (2001).

96. W. B. Hou, A. B. Kudryavtsev, T. L. Bray, L. J. DeLucas, and W. W. Wilson, *J. Cryst. Growth*, **232**, 265 (2001).
97. J. M. Garcia-Ruiz, *Methods Enzymol*,. **368**, 130 (2003).
98. C. N. Nanev, A. N. Penkova, and N. Chayen, *Ann. N. Y. Acad. Sci.*, **1027**, 1 (2004).
99. S. P. F. Humphreys-Owen, *Proc. R. Soc. Lond. A* **197**, 218 (1949).
100. K. Kurihara, S. Miyashita, G. Sazaki, T. Nakada, Y. Suzuki, and H. Komatsu, *J. Cryst. Growth*, **166**, 904 (1996).
101. A. McPherson, A. J. Alexander, Y. G. Kuznetsov, S. Koszelak, M. Wells, G. Jenkins, J. Howard, and G. Lawson, *J. Cryst. Growth*, 572 (1999).
102. D. Vivares, E. W. Kaler, and A. M. Lenhoff, *Cryst. Growth Des.*, **7**, 1411 (2007).
103. R. F. Sekerka, *J. Cryst. Growth*, **3/4**, 71 (1968).
104. W. F. Berg, *Proc. R. Soc. Lond. A*, **164**, 79 (1938).
105. C. N. Nanev, *Prog. Cryst. Growth Charact.*, **35**, 1 (1997).
106. C. N. Nanev and D. Iwanov, *Phys. Stat. Sol.*, **23**, 663 (1967).
107. D. E. Ovsienko, G. A. Alfintsev, and V. V. Maslov, *J. Cryst. Growth*, **26**, 233 (1974).
108. K. Onuma, K. Tsukamoto, and I. Sunagawa, *J. Cryst. Growth*, **98**, 384 (1989).
109. C. N. Nanev and R. S. Rashkov, *J. Cryst. Growth*, **158**, 136 (1996).
110. C. N. Nanev and A. N. Penkova, *J. Cryst. Growth*, **237–239**, 283 (2002).
111. C. N. Nanev, *Cryst. Res. Technol.*, **39**, 3 (2004).
112. A. E. S. Van Driessche,Universidad de Granada, 2007.
113. M. Sleutel, R. Willaert, L. Wyns, and D. Maes, *Cryst. Growth Des.*, **9**, 335 (2009).
114. A. A. Chernov, *J. Cryst. Growth*, **24/25**, 11 (1974).
115. C. N. Nanev, *J. Cryst. Growth*, **212**, 516 (2000).
116. M. Sato and M. Uwaha, *Phys. Rev. B*, **51**, 11172 (1995).
117. M. Uwaha, Y. Saito, and M. Sato, *J. Cryst. Growth*, **146**, 164 (1995).
118. S. R. Coriell, B. T. Murray, A. A. Chernov, and G. B. McFadden, *J. Cryst. Growth*, **169**, 773 (1996).
119. A. A. Chernov, S. R. Coriell, and B. T. Murray, *J. Cryst. Growth*, **132**, 405 (1993).
120. N. Cabrera and D. A. Vermileya, *Growth and Perfection of Crystals*, Wiley, New York, 1958.
121. J. P. van der Eerden and H. Müller-Krumbhaar, *Phys. Rev. Lett.*, **57**, 2431 (1986).
122. F. C. Frank, *Growth and Perfection of Crystals*, Wiley, New York, 1958.
123. O. Gliko, N. A. Booth, and P. G. Vekilov, *Acta Crystallogr. D*, **58**, 1622 (2002).
124. H. Lin, S. T. Yau, and P. G. Vekilov, *Phys. Rev. E*, **67**, 0316061 (2003).
125. G. Ehrlich and F. G. Hudda, *J. Chem. Phys.*, **44**, 1039 (1966).
126. R. L. Schwoebel and E. J. Shipsey, *J. Appl. Phys.*, **37**, 3682 (1966).
127. R. L. Schwoebel, *J. Appl. Phys.*, **40**, 614 (1969).
128. G. S. Bales and A. Zangwill, *Phys. Rev. B*, **41**, 5500 (1990).
129. N. Ming, K. Tsukamoto, I. Sunagawa, and A. A. Chernov, *J. Cryst. Growth*, **91**, 11 (1988).
130. H. Hondoh and T. Nakada, *J. Cryst. Growth*, **275**, e1423–e1429 (2005).
131. A. J. Malkin, T. A. Land, Y. G. Kuznetsov, A. McPherson, and J. J. DeYoreo, *Phys. Rev. Lett.*, **75**, 2778 (1995).
132. T. A. Land, A. J. Malkin, Y. G. Kuznetsov, A. McPherson, and J. J. De Yoreo, *Phys. Rev. Lett.*, **75**, 2774 (1995).

133. A. J. Malkin, Y. G. Kuznetsov, T. A. Land, and J. J. De Yoreo, *Nat. Struct. Biol.*, **2**, 956 (1995).
134. S. T. Yau and P. G. Vekilov, *Nature*, **406**, 494 (2000).
135. K. A. Jackson, *Growth and Perfection of Crystals*, Chapman and Hill, London, 1958.
136. A. E. Nielsen, *International Series of Monographs on Analytical Chemistry*, Vol. 1, Pergamon, New York, 1964.
137. J. D. Weeks, *Ordering in Strongly Fluctuating Condensed Matter Systems*, Plenum Press, New York, 1986.
138. M. Elwenspoek, *J. Phys. A Math. Gen.*, **20**, 669 (1987).
139. J. M. Garcia-Ruiz, *J. Struct. Biol.*, **142**, 22 (2003).
140. T. Hwa, M. Kardar, and M. Paczuski, *Phys. Rev. Lett.*, **66**, 441 (1990).
141. M. Elbaum, *Phys. Rev. Lett.*, **67**, 2982 (1991).
142. B. Mutaftschiev, *Handbook of Crystal Growth*, Vol. 1a, North-Holland, Amsterdam, 1993.
143. X.-Y. Liu, P. van Hoof, and P. Bennema, *Phys. Rev. Lett.*, **71**, 109 (1993).
144. K. Izumi, Gan Ping, M. Hashimoto, A. Toda, H. Miyaji, Y. Miyamoto, and Y. Nakagawa, *Advances in the Understanding of Crystal Growth Mechanisms*, North-Holland, Amsterdam, 1997.
145. R. A. Laudise, *J. Am. Chem. Soc.*, **81**, 562 (1959).
146. R. A. Laudise and A. A. Ballman, *J. Am. Chem. Soc.*, **81**, 2655 (1958).
147. R. A. Laudise and A. A. Ballman, *J. Phys. Chem.*, **64**, 688 (1960).
148. E. D. Kolb and R. A. Laudise, *J. Am. Ceram. Soc.*, **49**, 302 (1966).
149. L. A. M. J. Jetten, H. J. Human, P. Bennema, and J. P. van der Eerden, *J. Cryst. Growth*, **68**, 503 (1984).
150. X.-Y. Liu and P. Bennema, *J. Cryst. Growth*, **139**, 179 (1994).
151. S. D. Durbin and G. Feher, *J. Cryst. Growth*, **76**, 583 (1986).
152. S. Gorti, E. L. Forsythe, and M. L. Pusey, *Cryst. Growth Des.*, **4**, 691 (2004).
153. S. Gorti, E. L. Forsythe, and M. L. Pusey, *Cryst. Growth Des.*, **5**, 473 (2005).
154. E. v. Veenendaal, P. J. C. M. van Hoof, J. van Suchtelen, W. J. P. van Enckevort, and P. Bennema, *J. Cryst. Growth*, **198/199**, 22 (1999).
155. Y. Saito, *Statistical Physics of Crystal Growth*, World Scientific, Singapore, 1996.
156. J. F. Lutsko, V. Basios, G. Nicolis, J. Kozak, M. Sleutel, and D. Maes, *J. Chem. Phys.*, **132**, 035102 (2010).
157. B. Lorber, M. Skouri, J. P. Munch, and R. Giegé, *J. Cryst. Growth*, **128**, 1203 (1993).
158. M. Skouri, B. Lorber, R. Giegé, J.-P. Munch, and J. S. Candau, *J. Cryst. Growth*, **152**, 209 (1995).
159. J. Hirschler and J. C. Fontecilla-Camps, *J. Cryst. Growth*, **171**, 559 (1997).
160. R. A. Judge, E. L. Forsythe, and M. L. Pusey, *Biotechnol. Bioeng.*, **59**, 776 (1998).
161. G. Sazaki, S. Aoki, H. Ooshima, and J. Kato, *J. Cryst. Growth*, **139**, 95 (1994).
162. B. R. Thomas, D. Carter, and F. Rosenberger, *J. Cryst. Growth*, **187**, 499 (1998).
163. B. R. Thomas, P. G. Vekilov, and F. Rosenberger, *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 226 (1998).
164. T. Nakada, G. Sazaki, S. Miyashita, S. D. Durbin, and H. Komatsu, *J. Cryst. Growth*, **196**, 503 (1999).
165. M. W. Burke, R. Leardi, R. A. Judge, and M. L. Pusey, *Cryst. Growth Des.*, **1**, 333 (2001).

166. T. Matsui, G. Sazaki, H. Hondoh, Y. Matsuura, T. Nakada, and K. Nakajima, *J. Cryst. Growth*, **293**, 415 (2006).

167. C. Abergel, M. P. Nesa, and J. C. Fontecilla-Camps, *J. Cryst. Growth*, **110**, 11 (1991).

168. K. Provost and M. C. Robert, *J. Cryst. Growth*, **156**, 112 (1995).

169. J. Hirschler and J. C. Fontecilla-Camps, *Acta Crystallogr. D Biol. Crystallogr.*, **52**, 806 (1996).

170. F. L. Ewing, E. L. Forsythe, M. van der Woerd, and M. L. Pusey, *J. Cryst. Growth*, **160**, 389 (1996).

171. V. Bhamidi, B. L. Hanson, A. Edmundson, E. Skrzypczak-Jankun, and C. Schall, *J. Cryst. Growth*, **204**, 542 (1999).

172. C. L. Caylor, I. Dobrianov, S. G. Lemay, C. Kimmer, S. Kriminski, K. D. Finkelstein, W. Zipfel, W. W. Webb, B. R. Thomas, A. A. Chernov, and R. E. Thorne, *Proteins*, **36**, 270 (1999).

173. S. T. Yau, B. R. Thomas, O. Galkin, O. Gliko, and P. G. Vekilov, *Proteins*, **43**, 343 (2001).

174. I. Dobrianov, C. Caylor, S. G. Lemay, K. D. Finkelstein, and R. E. Thorne, *J. Cryst. Growth*, **196**, 511 (1999).

175. E. H. Snell, R. A. Judge, L. Crawford, E. L. Forsythe, M. L. Pusey, M. Sportiello, P. Todd, H. Bellamy, J. Lovelace, J. M. Cassanto, and G. E. O. Borgstahl, *Cryst. Growth Des.*, **1**, 151 (2001).

176. M. C. Robert, B. Capelle, B. Lorber, and R. Giegé, *J. Cryst. Growth*, **232**, 489 (2001).

177. I. Yoshizaki, A. Kadowaki, Y. Iimura, N. Igarashi, S. Yoda, and H. Komatsu, *J. Synchrotron Rad.*, **11**, 30 (2004).

178. A. Moreno, A. Theobald-Dietrich, B. Lorber, C. Sauter, and R. Giegé, *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 789 (2005).

179. I. Yoshizaki, S. Fukuyama, H. Koizumi, M. Tachibana, K. Kojima, Y. Matsuura, M. Tanaka, N. Igarashi, A. Kadowaki, L. Rong, S. Adachi, S. Yoda, and H. Komatsu, *J. Cryst. Growth*, **290**, 185 (2006).

180. A. E. S. Van Driessche, G. Sazaki, B. Guoliang, F. Otalora, J. A. Gavira, M. Takuro, Y. Izumi, K. Tsukamoto, and N. Kazuo, *Cryst. Growth Des.*, **9**, 3062 (2009).

181. B. R. Thomas, P. G. Vekilov, and F. Rosenberger, *Acta Crystallogr. D Biol. Crystallogr.*, **52**, 776 (1996).

182. S. D. Durbin and G. Feher, *J. Mol. Biol.*, **212**, 763 (1990).

183. H. Li, A. Nadarajah, and M. L. Pusey, *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 1036 (1999).

184. P. G. Vekilov and F. Rosenberger, *J. Cryst. Growth*, **186**, 251 (1998).

185. W. S. Cleveland, *J. Am. Stat. Assoc.*, **74**, 829 (1979).

186. Y. Iimura, I. Yoshizaki, H. Nakamura, S. Yoda, and H. Komatsu, *Jpn. J. Appl. Phys.*, **42**, 5831 (2003).

187. Y. Matsuura and A. A. Chernov, *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 1347 (2003).

188. G. Bliznakov, *Fortschr. Min.*, **36**, 149 (1958).

189. K. Sangwal, *Prog. Cryst. Growth Charact.*, **32**, 3 (1996).

190. J. J. De Yoreo, L. A. Zepeda-Ruiz, R. W. Friddle, S. R. Qiu, L. E. Wasylenki, A. A. Chernov, G. H. Gilmer, and P. M. Dove, *Cryst. Growth Des.*, **9**, 5135 (2009).

191. M. L. Weaver, S. R. Qiu, R. W. Friddle, W. H. Casey, and J. J. De Yoreo, *Cryst. Growth Des.*, **10**, 2954 (2010).

192. X. Y. Liu and K. Maiwa, *J. Chem. Phys.*, **106**, 1870 (1997).

193. A. E. S. Van Driessche, J. A. Gavira, and L. D. P. Lopez, *J. Cryst. Growth*, **311**, 3479 (2009).

194. Y. Suzuki, G. Sazaki, M. Matsumoto, M. Nagasawa, K. Nakajima, and K. Tamura, *Cryst. Growth Des.*, **9**, 4289 (2009).

195. A. E. S. Van Driessche, J. M. Garcia-Ruiz, J. M. Delgado-López, and G. Sazaki, *Cryst. Growth Des.*, **10**, 3909 (2010).

196. R. Wen, A. Lahiri, M. Azhagurajan, S. Kobayashi, and K. Itaya, *J. Am. Chem. Soc.*, **132**, 13657 (2010).

197. G. Sazaki, S. Zepeda, S. Nakatsubo, E. Yokoyama, and Y. Furukawa, *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 19702 (2010).

198. S. Yanagiya and N. Goto, *J. Cryst. Growth*, **312**, 3356 (2010).

199. G. Dai, G. Sazaki, T. Matsui, K. Tsukamoto, K. Nakajima, Q. Kang, and W. Hu, *Cryst. Growth Des.*, **11**, 88–92 (2011).

THE ABILITY OF SILICA TO INDUCE BIOMIMETIC CRYSTALLIZATION OF CALCIUM CARBONATE

MATTHIAS KELLERMEIER,[1,3] EMILIO MELERO-GARCÍA,[2]
WERNER KUNZ,[1] AND JUAN MANUEL GARCÍA-RUIZ[2]

[1]*Institute of Physical and Theoretical Chemistry, University of Regensburg, Universitätsstrasse 31, D-93040 Regensburg, Germany*
[2]*Laboratorio de Estudios Cristalográficos, IACT (CSIC-UGR), Avda. del Conocimiento s/n, P.T. Ciencias de la Salud, E-18100 Armilla, Spain*
[3]*Physical Chemistry, University of Konstanz, Universitätsstrasse 10, D-78457 Konstanz, Germany*

CONTENTS

Advances in Chemical Physics, Volume 151: Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanoscale Materials, First Edition. Edited by Gregoire Nicolis and Dominique Maes.

I. INTRODUCTION

The role of amorphous precursors in crystallization processes is a rapidly growing area of research and has attracted a considerable deal of attention in the past years. Recently, particular interest stems from the field of biomineralization, where amorphous mineral phases were identified both as temporary intermediates and in a permanently stable state as structural elements [1–4]. Striking examples for the latter case are the porous silica skeletons produced by various freshwater and marine algae, such as diatoms or radiolarians [5–7]. Their abundance in nature can readily be understood when considering the relatively high stability of amorphous silica against transformation to crystalline polymorphs [8]. By contrast, the amorphous phases of other widespread biominerals, such as calcium carbonate and calcium phosphate, are usually transient species that crystallize quickly in aqueous media [9, 10]. Nevertheless, frameworks comprising stable amorphous domains of both minerals were found in the tissues of diverse living organisms, including plant cystoliths [11], ascidians [12], calcareous sponges [13], or the cuticles of some crustaceans [14]. The advantage of using amorphous instead of crystalline material may lie in beneficial mechanical properties such as flexibility or in the commonly higher solubility of amorphous phases that allows intermittent storage and fast mobilization of calcium ions when needed (e.g., during moulting processes) [1–3, 15]. Beyond that, investigations of the formation mechanisms leading to the complex structures of selected biominerals have revealed that at least in some cases the inorganic component is precipitated first in an amorphous form to later crystallize in the particular polymorph that is most suitable for the given demands [16]. This was described in detail for the spicules, spines, and teeth of sea urchins [17–19] and also observed during the development of the shells of certain mollusks [20], the calciferous glands of earthworms [21], and the teeth of chitons [22], as well as in the mineralization of vertebrate bone [23] and dental enamel [24]. Though not clarified to date, the choice of a mineralization pathway involving amorphous intermediates might originate from the benefit that crystallization becomes in this way independent of solution parameters such as solubility products or pH [3, 25]. In addition, for lack of crystallographic symmetry restraints, amorphous phases can easily be shaped into any desired form before they crystallize, thus enabling the generation of complex crystal morphologies as confirmed in laboratory syntheses utilizing distinct artificial molds [26, 27]. These observations have stimulated a broad panoply of bioinspired crystallization assays meant to mimic strategies recognized in biomineralization in order to design novel materials with advanced structure and properties [28].

Calcium carbonate is often the chosen model system in this context not only due to its importance as abundant biomineral, but also in view of its rich polymorphism as well as its ample occurrence in geological sediments and corresponding implications for the natural CO_2 cycle and Earth's climate [29–31]. Further

fundamental relevance derives from its wide industrial application, for instance, as a filler or ingredient of construction materials [32], and the long-standing problems caused by incrustation and scaling due to spontaneous precipitation at elevated temperatures [33].

In the classical picture, crystallization is treated predominantly from a thermodynamic viewpoint and considered to take place by the formation of critical nuclei that subsequently grow through unit cell replication. However, such simple conceptualization often fails to explain distinct effects encountered in many practical systems, and there appear to be other reaction channels that account in particular for the complexity of biominerals and related synthetic materials [34]. For instance, it was demonstrated only recently that neutral ion clusters occur in solutions of calcium carbonate prior to nucleation [35]. These so-called prenucleation clusters (PNC) proved to be stable in equilibrium with the dissolved ions and therefore differ fundamentally from the critical clusters envisaged by classical theories that are thought to be metastable and a rare species as they result from statistical fluctuations. This finding delineates entirely new potential scenarios for nucleation that are currently under debate. Another circumstance complicating the situation is the fact that, especially for a polymorphic compound such as calcium carbonate, there exists an alternative pathway for crystallization controlled by kinetics rather than thermodynamics. In this case, the mineral does not crystallize directly in its thermodynamically stable modification (calcite in the case of $CaCO_3$), but may undergo multiple phase transitions before, along a series of intermediates toward increasing stability [36]. This behavior is reflected in Ostwald's rule of stages, which states that during sequential precipitation, the least stable and most soluble polymorph is formed first and, in succession, converted preferentially to phases that are structurally and energetically closest [37]. For $CaCO_3$, experimental studies have confirmed that kinetically driven crystallization is particularly promoted in systems at high supersaturation and initially affords nanoparticles of amorphous calcium carbonate (ACC) [38, 39]. This is in line with Ostwald's law, which would predict a kinetic phase sequence according to ACC $\rightarrow$ vaterite $\rightarrow$ aragonite $\rightarrow$ calcite on the basis of relative stabilities under ambient conditions, as shown in Fig. 1. In fact, freshly precipitated ACC transforms to calcite rather rapidly when left in contact with the solution, either directly or involving vaterite as metastable crystalline intermediate [40]. Aragonite is usually not observed at room temperature but replaces vaterite in the sequence upon heating [41]. Thus, the determination of whether crystallization takes place under kinetic control and which of the metastable polymorphs occur in such a multistep scenario essentially depends on the chosen experimental parameters.

Another means to influence the process is to introduce additives that will interact with one or more of the kinetic intermediates and hence potentially affect their interconversions. For example, there have been numerous attempts to prevent ACC from transforming to crystalline polymorphs *in vitro*, partly for the sake of

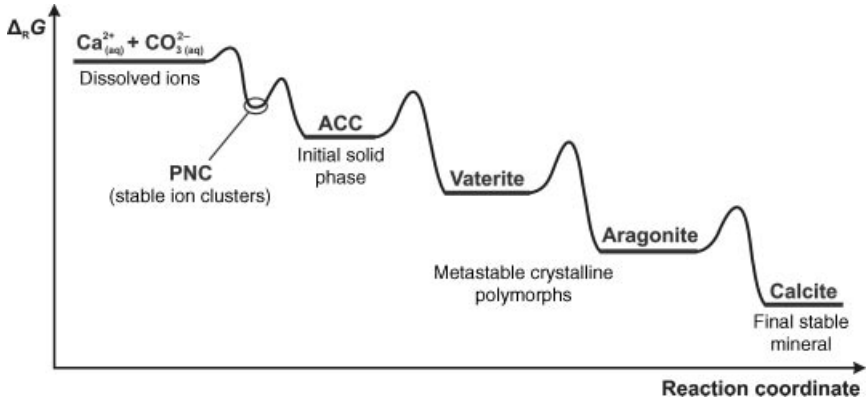


Figure 1. Crystallization of calcium carbonate under kinetic control, according to Ostwald's rule of stages and considering the existence of ion clusters (PNC) in a minimum of Gibbs energy prior to nucleation. The indicated relative stabilities apply to moderate temperatures and pressures. Note that not all the possible precursor phases necessarily occur in such a sequential scenario toward the final stable mineral calcite. Often, either vaterite or aragonite is observed as metastable intermediate, whereas the formation of ACC as the first solid phase upon nucleation has shown to be common over a wide range of conditions. The scheme is partly based on drawings given in Refs 35, 36.

scale inhibition [42]. Substantially prolonged lifetimes of ACC have been realized, among others, by adding foreign ions such as Mg^{2+} [43] or phosphate [39], and by the use of specific organic inhibitors such as certain surfactants [44], polymers [45], dendrimers [46], or phytic acid [47]. Increased kinetic stability is thereby often achieved by adsorption of additive species on the surface of the metastable particles, which impedes dissolution and thus hampers common solution-mediated transformation routes. On the same basis, vaterite and aragonite also could be stabilized against energetically favored conversion to calcite [48, 49]. Nevertheless, with regard to the above-described emerging body of knowledge on amorphous intermediates in biomineralization, it seems of particular interest to temporarily stabilize ACC in the course of $CaCO_3$ precipitation and employ it as a precursor for the construction of elaborate crystalline architectures. Indeed, evidence for such a crystallization pathway has been reported frequently and for a variety of materials in biomimetic approaches of morphosynthesis [50]. In analogy to phenomena believed to occur *in vivo*, surfactant- or polymer-protected amorphous nanoparticles were found to assemble into networks acting as depots that supply mineral units only gradually and hence uncouple morphogenesis from particular solution conditions. Under suitable circumstances, this may excite self-organization of small crystal building blocks into distinct superstructures [51–54], often through mesoscale transformations via a nonclassical route of crystallization [34, 36].

In the rich pool of additives applied for mineralization purposes to date, dissolved silica takes a fairly exceptional position as it is an inorganic compound and nevertheless capable of inducing delicate effects comparable to what can be achieved even by most complicated organic counterparts. Already around 30 years ago, it was discovered that precipitation of alkaline earth carbonates in silica-rich media at elevated pH can afford crystal architectures with unusual shapes and structures [55]. Under ambient conditions, hierarchical "sheaf-of-wheat" patterns consisting of cleaved calcite rhombohedra were obtained with $CaCO_3$ [56], while the higher homologues $SrCO_3$ and $BaCO_3$ self-assembled spontaneously into aggregates of uniform elongated nanocrystals that displayed stunning non-crystallographic morphologies such as sinuous sheets and twisted filaments [57–59]. Coral-like polycrystalline particles of aragonite exhibiting a similar mode of construction were later grown at elevated temperature or by adding seed crystals [60, 61]. Owing to their striking resemblance to biogenic minerals, these materials were named "silica biomorphs." They exemplify that curved shapes and higher order textures—intuitively associated with the animate nature—may very well be produced also by the interplay of purely inorganic species, thus disqualifying morphology as a tool to identify biogenicity [62–64].

A scenario accounting for morphogenesis of complex form in these simple systems has recently been introduced [65, 66]. It was proposed that the driving force governing self-assembly relies on a coupling of the chemistry of the components in aqueous solution. In this paradigm, the silica does not behave as a true soluble additive, but rather interferes with carbonate crystallization by factual coprecipitation. Since the speciation and solubility of both carbonate and silicate are essentially determined by the pH of the medium, their precipitation cannot be considered independent and a distinct mutual influence on short length scales is to be expected under certain conditions. More precisely speaking, carbonate particles growing in moderately alkaline environments provoke a sensible pH decrease in their vicinity due to the release of protons upon dissociation of bicarbonate ions existing in equilibrium. This local change in pH is experienced by silicate species and triggers their polymerization. Eventually, a skin of amorphous silica is formed around the carbonate crystallites, which thus become stabilized at nanometric dimensions and represent the building blocks from which crystal aggregates characteristic of silica biomorphs are constructed [65, 66].

Inspired by these findings, we have started investigations meant to elucidate the potential roles of added silica during polymorphic precipitation of calcium carbonate on a fundamental level. In this context, interactions between silicate species and $CaCO_3$ precursors occurring in kinetically controlled crystallization were in the focus of interest, rather than deliberately aiming at the synthesis of biomorphic materials. Corresponding experiments have shown that silica is a highly versatile additive able to substantially enhance the endurance of any metastable intermediate if certain parameters are adjusted properly. Furthermore, a pronounced influence

was also verified for ion clusters in the prenucleation regime and in particular their transformation to nanoparticles of ACC. Thus, the addition of silica allows arresting the crystallization process at virtually any of the stages shown in Fig. 1. In this manner, profound insight into the mechanisms of nucleation and subsequent phase interconversion could be gained, and strategies toward a concerted stabilization and isolation of different polymorphs were developed.

The present contribution highlights central aspects of our work concerning the effect of silica on the initially nucleated amorphous phase. In a recent study, we have demonstrated that the pH-based chemical coupling assumed to drive the formation of silica biomorphs can be applied to coat ACC nanoparticles with distinct layers of silica during precipitation from solution [67]. These shells were found to be porous and permeable to reagents at low silica concentrations. Under these conditions, crystallization took place after considerable periods of delay and at markedly decreased rates. By contrast, higher amounts of the additive caused dense cementation of the ACC grains in a silica matrix and thus enabled permanent conservation of the amorphous precursors on timescales of years. Here, we extend this concept and describe how silica-mediated stabilization of ACC can contribute to the morphogenesis of complex crystal forms in such environments, thus facilitating structured crystallization even at nominally high supersaturation. To that end, calcium carbonate was precipitated by adding calcium chloride to mixtures of sodium carbonate and sodium silicate. The evolution of samples with varying silica content was followed with time and the as-formed ACC particles, their gradual transformation to calcite, and the morphology and structure of the final crystalline products were characterized by different techniques. The observed growth behavior, involving intermittently stabilized ACC, bears fundamental analogy to apparent principles in biomineralization and modern morphosynthesis, and hence underlines the biomimetic character of these plain inorganic precipitation systems.

II. EXPERIMENTAL SECTION

A. Crystallization Experiments

First, suitable sodium silicate solutions were prepared by diluting commercial water glass stock (Sigma-Aldrich, reagent grade, containing 26.7 wt% SiO_2 and 13.8 wt% NaOH) in ratios ranging from 1:250 to 1:1400 (v/v) with water of Milli-Q quality. These sols were purged with H_2O saturated nitrogen for about 1 h to dispel previously absorbed carbon dioxide and then used to dissolve 10 mM Na_2CO_3 (Roth, anhydrous, $\geq$99%). The resulting solutions were filtered (450 nm), stored in tightly stoppered plastic bottles, and replaced by fresh ones at least weekly. To trigger precipitation of calcium carbonate, 5 mL silica–carbonate mixture was filled in polypropylene vessels and the same volume of a CO_2-free 10 mM solution

of $CaCl_2 \cdot 2H_2O$ (Riedel-de Haën, ACS reagent, $\geq$99%) was added at once with the aid of a graded pipette ($t = 0$). The vials were closed and samples were allowed to age under quiescent conditions at $20 \pm 1^{\circ}C$. Thus, final species concentrations were 5 mM $CaCO_3$ and 0–750 ppm SiO_2. The starting pH was confirmed to be constant at 10.6 ± 0.1 over the entire studied range of silica contents, and there were no significant changes with time when silica was present due to its buffering ability. At the given high supersaturation, instant homogeneous nucleation of ACC nanoparticles imparted a cloudy appearance to the samples straight after mixing. In the reference system without silica, these particles usually remain suspended in solution until transformation to crystalline polymorphs is completed. Addition of silica in turn led to partial flocculation and, after some delay, deposition of an initially amorphous precipitate on the bottom of the vials, the amount of which increased with the silica concentration. Crystalline structures developed in all samples and were observed both adhering firmly to the vessel walls and growing directly on or within the amorphous floccules, in particular at higher silica contents. The time after mixing at which first crystals were sighted was found to be the longer the more silica was added. Quenching of the reaction at distinct stages and isolation of the formed precipitates were accomplished in two different ways. On the one hand, the suspensions were passed through membrane filters (Whatman, 200 nm) together with the sedimented floccules and potentially occurring crystals. The latter were detached from the walls by careful manipulation with a fine brush and transferred onto the filter by rinsing with a small quantity of water. The extracted solid was immediately frozen in liquid nitrogen and subsequently lyophilized for several hours using an Edwards Freeze Dryer Modulyo, yielding a white powder. Alternatively, to separate crystalline products from amorphous particles, the mother liquor was withdrawn and replaced by water. The flocculated material was then dispersed vigorously such that the crystals were released from the particle networks and settled to the ground. Suspended floccules could thus be selectively removed along with the supernatant by sucking with a pipette. This procedure was repeated until no more amorphous precipitates could be distinguished. Finally, the residual crystals were washed twice with water and ethanol and left to dry in air.

B. Analytical Methods

1. Visual Observations

The evolution of the samples with time was followed visually by inspecting the vials in regular intervals until ongoing crystallization processes were apparently terminated, that is, when either all amorphous particles had vanished and a clear solution with crystals on the walls was achieved or, at higher silica concentrations, when the amount of precipitated floccules did no longer change noticeably. Complementarily, a series of photographs were taken from the solutions over periods up

to months. For this purpose, experiments were carried out in plain glass tubes. Pictures were taken with a Canon PowerShot A640 digital camera connected to a PC.

2. *Light Microscopy*

Crystalline architectures obtained from the syntheses were investigated first with an optical microscope (Nikon Eclipse E400). An attached Canon EOS 350D camera was used for imaging. Coagulated particle networks as well as crystals associated with or embedded in them were studied by decanting the bulk volume of the supernatant solution at defined times after mixing and loading parts of the floccules together with the remaining mother liquid on glass slides. Comparative analyses showed that this treatment did not affect the state of the samples and there was no induced additional conversion of ACC species to crystalline polymorphs upon manipulation. The observed structures can thus be considered representative of the situation in solution at the corresponding time.

3. *Electron Microscopy*

Samples of the suspended nanoparticles and those in the precipitated floccules were prepared for transmission electron microscopy (TEM) by placing aliquots from the top solution and the bottom region of the vials, respectively, on carbon filmed copper grids (200 mesh, Plano GmbH) that were hydrophilized previously by glow discharge. After soaking for 1 min, the liquid was removed with a filter paper. The adsorbed particles were cleaned of excess salt by spreading and rapidly blotting a drop of water on the grids, followed by drying in air. This technique allows an effective quenching of the reaction and there are no severe structural changes during isolation, given that identical results were obtained when freezing the solutions and studying particles in a vitrified state by cryo-TEM [67]. The crystal aggregates formed at higher silica concentrations were examined by gathering a practicable amount of sample, crushing the specimens, and grinding them in ethanol to yield a fine suspension, a small volume of which was subsequently evaporated on a grid. TEM analyses were performed on a Philips CM 12 microscope at 120 kV under low-dose conditions, using CCD cameras from Gatan (model TV 673) and TVIPS (model TEM1000). Images were acquired and processed with the EM-MENU 4 software package.

For field emission scanning electron microscopy (FESEM), both the freeze-dried powders and separated crystals were transferred onto conducting double-sided adhesive carbon tabs that were fixed on standard aluminum SEM pin stubs. Samples were coated with carbon prior to analysis using a Balzers MED 010 sputter coater. FESEM studies were carried out on a Zeiss LEO Gemini 1530 microscope at an acceleration voltage of 2 or 3 kV and working distances in the range of 3–5 mm. Most images were recorded by mixing signal from the in-lens and secondary

electron detectors. EDX measurements were conducted at 10 kV with the aid of an installed Oxford INCA microanalysis system. Spectra from amorphous particle networks were collected from large areas to gain the mean composition of numerous individuals at once. In the case of the crystalline products, several typical specimens were probed independently at each silica concentration, sampling a series of different positions per particle and averaging results for a given sample.

4. Turbidity Measurements

The duration of crystallization was estimated by monitoring the turbidity of the solutions as a function of time. Experiments were performed using an automated setup built in-house [68]. For measurement, reagents were mixed directly in corresponding cuvettes and inserted quickly into the apparatus. Data were acquired while stirring samples, reading values approximately every 17 s. Clearing times were determined from the inflection point of the sharp decrease in the turbidity–time profiles seen upon completion of crystallization.

5. X-ray Diffraction

The polymorphism of the isolated crystalline objects was investigated by picking up individuals using a microloop wetted with a drop of oil. Diffraction of the specimens was measured with a Bruker X8 Proteum setup (Kappa configuration) comprising a Microstar high brilliance rotating anode source (Cu-K_α radiation) and a Smart 6000 two-dimensional CCD detector having 512×512 pixels. During data collection, samples were rotated around the φ-axis by 180° in steps of 5°, the integration time per image being 20 s.

III. RESULTS

A. Stabilization of ACC in the Presence of Silica

Upon mixing 10 mM aqueous solutions of calcium chloride and sodium carbonate, fast precipitation of calcium carbonate instantly yields an amorphous phase consisting of fairly uniform spherical nanoparticles with typical diameters around 150 nm [67]. In the absence of any further substances, these particles persist in solution for a certain time depending on specific parameters such as temperature or the detailed way of mixing, but eventually and usually after no longer than a few hours transform quantitatively into more stable crystalline polymorphs such as vaterite or calcite. Analyses of corresponding particles formed in alkaline systems containing sodium silicate show that under these conditions, the ACC grains become spontaneously enveloped by continuous skins of amorphous silica (see Fig. 2a and b). The thickness of these outer layers, being in the range of 1–2 nm at the lowest studied concentration (135 ppm), increases with the amount of silica in

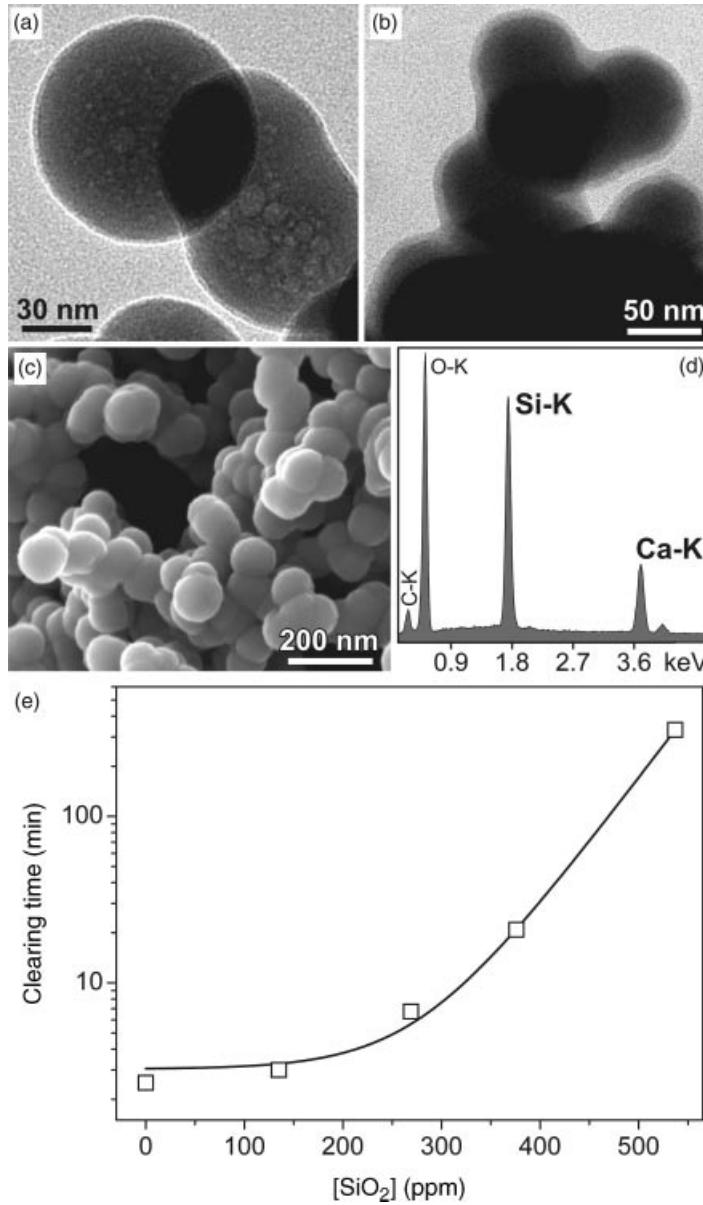


Figure 2. Deposition of silica on growing ACC particles upon precipitation of $CaCO_3$ in alkaline media. (**a**) TEM image of nanoparticles suspended in solutions at 270 ppm SiO_2, isolated 1 min after mixing. The dark blistering carbonate core is covered by a lighter thin layer of amorphous silica.

solution to reach values of about 10 nm at 750 ppm SiO_2. Deposition of silica on the ACC nanoparticles thus produces core–shell composite materials with silica contents of up to 40 wt% SiO_2, as evidenced by EDX measurements (Fig. 2d). In turn, the mean size of single grains decreases as more silica is added, since sheathing of the carbonate particles by silica prevents further growth and the formation of the skins is faster at higher silica concentrations. Apart from that, the presence of silica all over the surface of the ACC precursors allows condensation reactions ($2\ -Si–OH \rightarrow\ -Si–O–Si-\ +\ H_2O$) in between the particles, facilitating cross-linking and conjunction of individuals. As a consequence, networks of densely agglomerated and partly fused silica-coated ACC nanoparticles are formed (Fig. 2c), which precipitate from solution and sediment as amorphous floccules to the bottom of the vial.

The effect of this spontaneous coating on the temporal progress of $CaCO_3$ crystallization was characterized by tracing the turbidity of the samples. Since the amorphous nanoparticles remain to a certain extent suspended and therefore becloud the mixtures, the time elapsed before the solution clears off is a direct measure for the effective duration of crystallization, given that crystals grow only nearby or attached to vessel walls. The detected clearing times were found to increase exponentially with the silica concentration (Fig. 2e). Total conversion of ACC required more than 5 h in the presence of 540 ppm SiO_2, while clearing of the solution already occurred after 2.5 min in the absence of silica. This implies that added silica decelerates the crystallization process.

Visual observations qualitatively confirm the trend indicated by the turbidity data. In this case, completion of crystallization can be equalized not only to a clearing of the solutions, but also and in particular at higher silica concentrations to the disappearance of the coagulated amorphous material. Inspecting the samples suggests that the floccules dissolve gradually until finally only discrete crystalline objects, adhered to the vial walls, are left. We note that the turbidity measurements had to be performed under moderate stirring to ensure reproducibility of the data, in contrast to the quiescent conditions chosen otherwise in this work, thus rendering results from the two methods not directly comparable. Durations determined

←

(**b**) Micrograph showing the border region of the flocculated material at 750 ppm SiO_2 after 90 min. The silica shells around the ACC particles are much more distinct than at lower concentrations and frequently span over multiple individuals. (**c**) High-magnification FESEM image of the floccules at 750 ppm SiO_2, evidencing that the presence of silica promotes cross-linking of the ACC grains and leads to the formation of three-dimensional networks. (**d**) EDX spectrum of the particles shown in (c), from which a Si/Ca atomic ratio of 1.57 ± 0.04 was calculated. (**e**) Semilogarithmic plot of the clearing time determined by turbidity measurements versus the amount of added silica, with a tentative exponential fit (full line) of the experimental data (open squares) ($t_c = a_0 + a_1 \cdot \exp([SiO_2]/a_2)$ with $a_0 = 3.0 \pm 0.6$ min, $a_1 = 0.021 \pm 0.004$ min, $a_2 = 55 \pm 2$ ppm, and $R^2 = 0.99998$).

visually for stagnant solutions are considerably longer than those inferred from the turbidity curves. Values range from about 70 min in the reference experiment without silica, over roughly 1 day at 375 ppm SiO_2, up to several months at 540 ppm. For the highest investigated silica concentration (750 ppm), both techniques have verified that only a certain fraction of the initially generated ACC–silica particles (around 50% [67]) dissolves upon aging, while the rest remains stable in the long term regardless of whether the samples are stirred or not.

These findings demonstrate that addition of silica progressively impedes the formation of crystalline $CaCO_3$ polymorphs from amorphous precursors. This inhibition can be ascribed to the silica skins precipitated around growing ACC grains soon after mixing. They restrain dissolution of the ACC core and prevent direct contact of particles necessary for possible solid-state conversions, consequently protecting the actually transient phase against ripening and energetically favored transformation [67]. Interestingly, a stabilizing effect of silicate ions on ACC was observed independently by Gal et al., who examined the thermal stability of both synthetic and biogenic ACC powders containing varying amounts of silica [69]. Their results evidenced that the temperature required for crystallization of calcite to occur increases with the silica content of the ACC samples. Based thereon, and in light of the abundance of silica in many plants [5–7], it was suggested that silicate species could perform the function of a stabilizing agent for ACC in biomineralization processes, apart from forming mineral deposits in tissues by themselves. The present data support this hypothesis and furthermore provide a mechanism accounting for silica-induced stabilization of ACC.

B. Crystallization of $CaCO_3$ from Silica-Coated ACC

The development of crystalline structures in the systems can readily be monitored by stopping the reaction at different times and analyzing the existing precipitates by means of SEM and TEM. In general, data show that the onset of crystallization becomes increasingly delayed as more silica is added. Coherently, the period up to which amorphous particles were discerned in the samples was longer at higher silica concentrations, in line with the conclusions drawn from turbidity measurements. At 270 ppm SiO_2, first crystals were sighted around 30 min after mixing. Subsequently, they grew at the expense of the nanoparticles until after about 3 h all amorphous material had vanished and only micron-sized crystalline architectures were left. In most cases, crystallization took place in the immediate vicinity of coagulated floccules, and crystals were often found to evolve directly on top of the networks, with a multitude of precursor particles decorating or tightly adhering to crystal faces (Fig. 3a–c). This suggests in principle that transformation of ACC occurs via a solid-state restructuring pathway. However, since the carbonate nanoparticles are shielded against each other and the growing crystal surface by their silica skins, this is not possible and the only feasible way of ACC conversion to crystalline phases is dissolution of the $CaCO_3$ core through the outer shell and subsequent

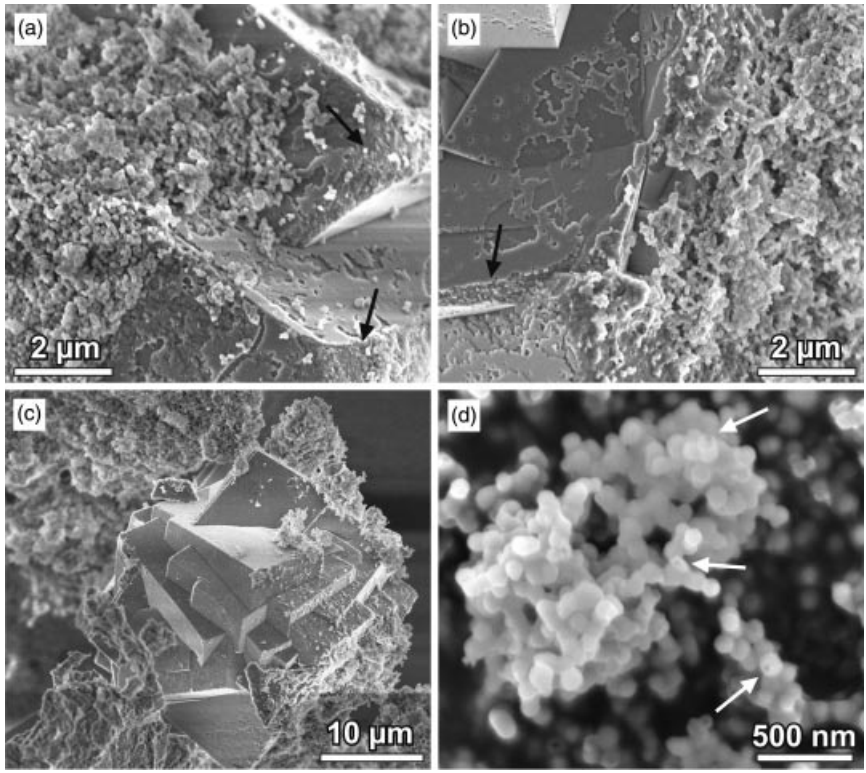


Figure 3. Transformation of silica-stabilized ACC to calcite at 270 ppm SiO_2. (**a**–**c**) FESEM images of samples quenched 1 h after mixing, revealing that crystals emerge from networks of amorphous nanoparticles and grow attached to the floccules in the following. Black arrows indicate sites where truncation of corners or edges of the rhombohedral calcite habit is obvious. (**d**) Closeup view of particle agglomerates isolated after 30 min. White arrows point to individuals that appear to have holes or cracks as a consequence of ongoing disintegration.

reprecipitation. This is supported by high-magnification SEM images of the amorphous particles during crystallization that, in some instances, disclose that the surface of single grains is perforated and their interior seems to be hollow (Fig. 3d). In this manner, the silica-coated ACC particles act as depots that constantly supply the surrounding solution with growth units. Consequently, the supersaturation is expected to be highest near the precipitated floccules, thus explaining why crystals are preferentially nucleated upon or inside the networks. The rate of $CaCO_3$ release from the core of the composite particles is essentially determined by the amount of silica added. Higher concentrations lead to the formation of less porous skins [67], which slow down the transport of reagents and hence accretively enlarge the window of time during which crystallization is active. At silica contents of up to

540 ppm, all amorphous particles are eventually consumed and, remarkably, the silica shells also dissolve after having released their ACC core. Therefore, the concentration of silica in solution should increase continuously as crystallization proceeds.

Similar observations were made as well for samples drawn from mixtures containing 750 ppm SiO_2. Figure 4 gives a gallery of typical images visualizing

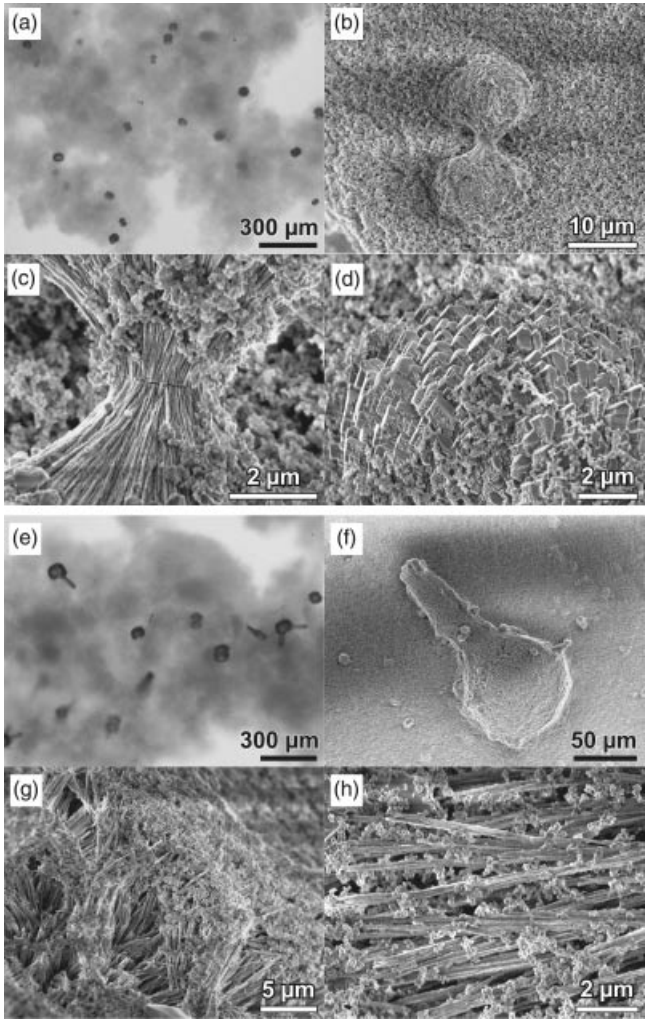


Figure 4. Optical and FESEM micrographs illustrating the growth of peculiar crystal aggregates from heavily condensed meshes of silica-coated ACC particles in samples at 750 ppm SiO_2. Precipitates were isolated after (**a**–**d**) 12 h and (**e**–**h**) 1 week.

the evolution of crystalline matter in the system. In this concentration regime, the amount of flocculated material is fairly large such that spacious three-dimensional networks extending over the entire bottom of the vials are generated (cf. Fig. 4a and e). Transformation of ACC sets in after around 5 h to produce rounded and dumbbell-shaped structures with diameters of 10–20 μm in the following 7 h (Fig. 4a and b). Growth of these forms continues up to about 1 day after mixing, while subsequently no further changes, neither in the size and morphology of particles nor in the overall composition and crystallinity of the samples, could be distinguished. Micrographs taken from precipitates isolated after standing for 1 week confirm that a bulk part of the originally deposited amorphous floccules still persists at this stage, along with numerous bigger spherical objects (or notched spheres), from which frequently thick linear strands sprout outward (Fig. 4e and f). These uncommon morphologies usually develop using beds of aggregated amorphous nanoparticles as substrates, in which they appear to be partially buried (Fig. 4b and f).

Zooming into the as-grown dumbbell-like and spheroidal architectures reveals that both are constituted by fibrous crystallites (Fig. 4c, g, and h). Often, clusters of ACC–silica precursor particles are seen to be firmly associated with needle-shaped individuals (Fig. 4c), decorate the crystal aggregates in the form of extended superficial layers (Fig. 4g), and sometimes even interpenetrate the assembly (Fig. 4h). TEM analyses of mature samples prove the presence of condensed spherical nanoparticles next to a multitude of uniform elongate crystals. Furthermore, frameworks of hollow particles were repeatedly observed adjacent to aggregates of crystallites (Fig. 5), in particular during the earlier stages of crystallization. EDX data demonstrate that these remnants consist of only silica and hence suggest that the ACC core but not as yet the silica skin was dissolved. Again, this is clear evidence that silica-coated ACC transforms into crystalline polymorphs through dissolution–renucleation processes. Those particles that endure in solutions at 750 ppm SiO_2 upon prolonged aging in turn seem to be sheathed by layers of somewhat lower porosity, which efficiently prevent exchange between the enclosed volume and the surrounding medium.

C. Morphology and Texture of the Final Crystalline Products

The vast majority of crystals isolated at the end of experiments performed without added silica were regular rhombohedral characteristic of calcite. The perfection of the rhombohedral shape recognized in most cases indicates that the crystalline phase grows via a dissolution-assisted route also in the silica-free reference [70]. Sporadically, spherical morphologies were sighted next to the euhedral forms, indicating that the samples contained traces of vaterite. The presence of 135 ppm silica was found to have no noticeable effect on the appearance of the final crystals (Fig. 6a). Indeed, the surface of the rhombohedra happens to be covered by a film of silica made of densely agglomerated colloidal spheres. This feature may however

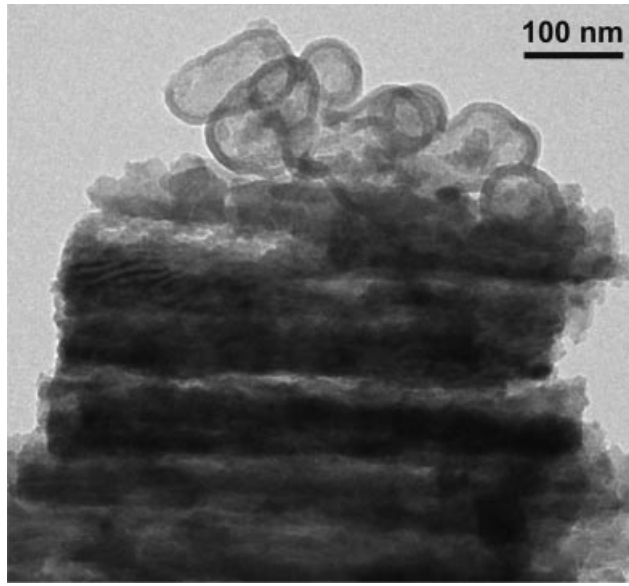


Figure 5. TEM image of structures formed in solutions containing 750 ppm SiO_2 after aging for 1 week, with hollow silica particles lying on top of a stack of piled rod-shaped crystallites.

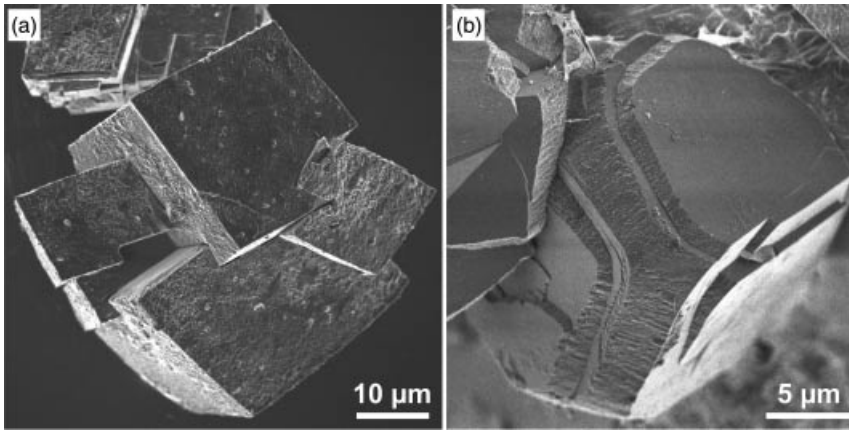


Figure 6. Final products of $CaCO_3$ crystallization conducted in the presence of (**a**) 135 and (**b**) 270 ppm SiO_2. Crystals were isolated after a period of 24 h, that is, well after all amorphous precursor particles had vanished. Classical rhombohedra with colloidal silica adsorbed unspecifically on their surfaces are observed at 135 ppm SiO_2, while characteristic rounding of corners and edges occurs at 270 ppm.

originate from secondary deposition of silica on the crystals upon isolation and drying. On raising its concentration, the silica not only stabilizes ACC precursors and thereby delay crystallization, but also starts to distinctly affect crystal growth and the emerging morphologies. At 270 ppm SiO_2, edges and corners of the classical rhombohedral habit become truncated especially in the later course of the crystallization process and are thus replaced by oblique, often rough faces (Figs. 6b and 3a and b). We note that at silica contents of both 135 and 270 ppm, the formation of conglomerates of intergrown crystals seems to be favored (cf. Figures 2c and 5a), whereas in the control experiment at 0 ppm predominantly individual rhombohedra were obtained. This is reasonable when considering that the ACC–silica particle networks are preferential sites for nucleation. Therefore, it is likely that multiple events of nucleation will take place close to one another. Continued crystal growth then inevitably leads to merging and the pronounced occurrence of intergrown architectures.

Increasing the amount of silica to 375 ppm brings about further profound morphological and structural changes. After equilibration and quantitative consumption of all ACC materials, a relatively small number of quite large, isolated crystalline objects were recovered from the bottom and the walls of the vials, which had adopted exceptional forms as shown in Fig. 7. From a phenomenological point of view, the observed morphologies evolve at first by dilation of an initial regular rhombohedron through extrusion of multiple fibrous projections with diameters in the range of 150–300 nm (Fig. 7a). The resulting elongated crystal then grows preferentially at its two outer tip regions, where the projections develop platy, wing-like domains and hence impart a dumbbell-shaped appearance to the specimen (Fig. 7b). Remarkably, each strand consists of three wings arranged at angles of $\sim 120^\circ$ between one another, such that the cross section of the strand takes the form of a three-pointed star (see inset in Fig. 7c). From these symmetry aspects, we infer that the long axis of the projections corresponds to the crystallographic *c*-axis of calcite, which is the only ternary axis in the rhombohedral lattice. Individual wings reach widths of up to while being as thin as 100–200 nm, and often exhibit a somewhat ribbed surface texture. Upon continued growth, such characteristic projections eventually sprout in all directions from the central crystal, leading to spherulitic entities decorated uniformly by three-pointed stars (Fig. 7c).

Crystallization of calcium carbonate in solutions with higher silica content yields in the formation of further intricate morphologies that are, especially in terms of their mode of construction, substantially different from those produced at 375 ppm (Fig. 8). Though forming on distinct timescales and consuming to varying degrees the available ACC precursors, the growth behavior and final structure of the crystalline products obtained at 540 and 750 ppm SiO_2 are widely identical. Once initiated, the morphological evolution in these systems can be summarized as follows. First, elongated seeds are discerned within the precipitated amorphous floccules, which transform consecutively into dumbbell shapes (Figs. 8a and 4b

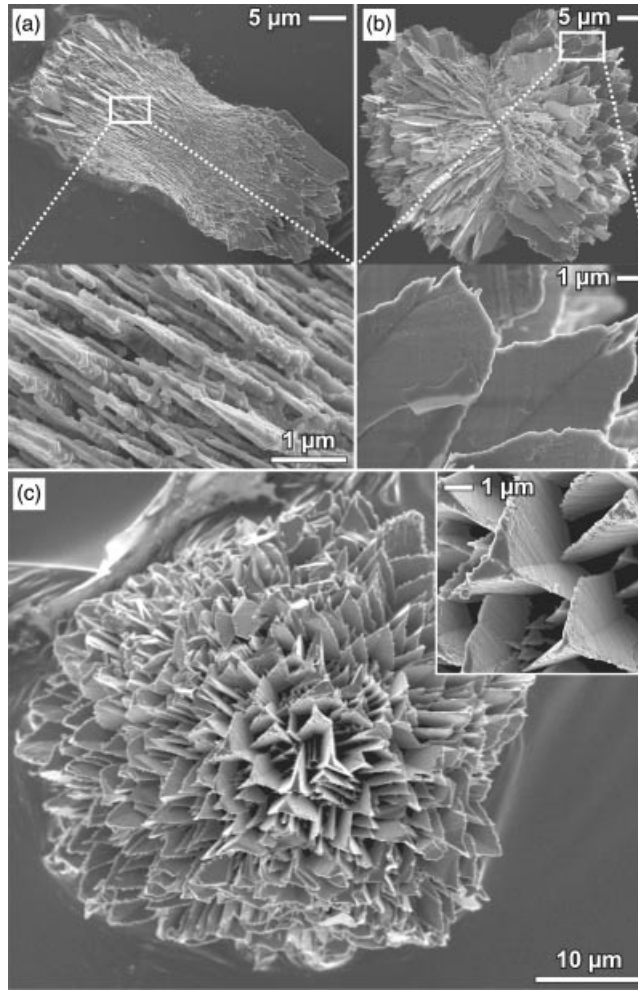


Figure 7. FESEM micrographs of crystal architectures produced at 375 ppm SiO_2 after 24 h.

and c) and, later, closed spherulites with diameters of around 100 µm (Fig. 8b). The latter are porous, with manifold voids characterizing their surface that lend a flowery appearance to the architectures. In the final step, usually after some days, rather thick trunk-like outgrowths develop from the spherulites (Figs. 8c and 3f), such that the resulting morphologies to some extent resemble those of natural radiolarians [5]. Mostly, one or two of these extrusions sprout from a given spherulite and grow in a quasilinear fashion over lengths of up to 100 µm. In line with the observations made for crystals isolated while still embedded in networks

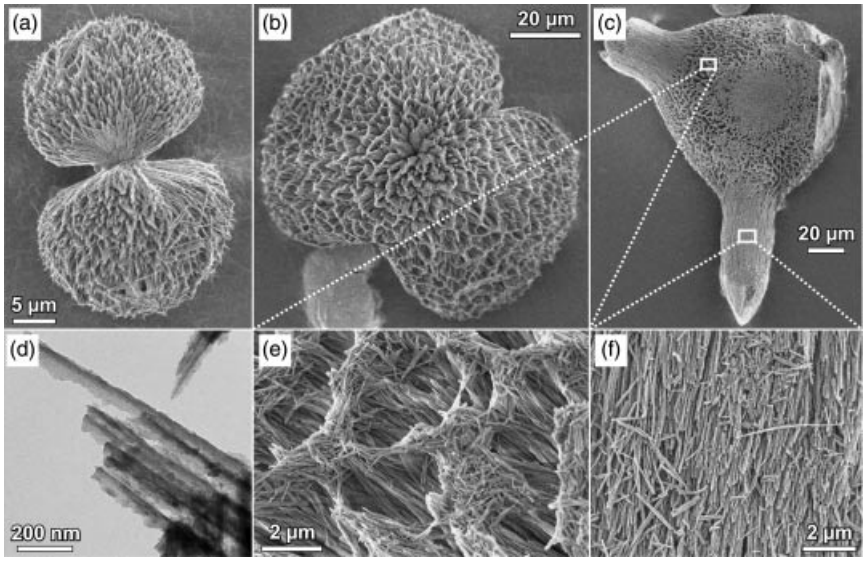


Figure 8. Crystal aggregates obtained at a silica concentration of 750 ppm. (**a**) A dumbbell-shaped particle isolated after 12 h. (**b**) Porous spherulitic architecture collected after 1 day. (**c**) After 1 week, trunk-like outgrowths have emerged from the spherulites. (**d**) TEM micrograph of a crushed sample grown for 1 week, showing aligned needle-like microcrystallites. (**e** and **f**) Zooms into the aggregate shown in (c), spotting the voids in the central spheroidal part and the co-orientation of the crystallites within a trunk. Note that the specimens were cleaned from amorphous material prior to isolation (in contrast to the samples depicted in Fig. 4).

of amorphous nanoparticles (cf. Fig. 4c, g and h), the precipitates were at all growth stages confirmed to be aggregates of myriad acicular crystallites exhibiting a fairly narrow size distribution (Fig. 8d–f). The diameter of the needles was found to be more or less constant throughout the entire assembly, usually measuring between 70 and 120 nm. By contrast, their length seems to vary depending on the position within the aggregate; that is, individuals become shorter as growth proceeds. Typical values range in a quite broad interval of about 500 nm to 5 µm. Generally, the needles tend to be oriented with their long axis pointing along the respective current growth direction. Therefore, neighboring crystallites are aligned preferentially parallel to each other (cf. Fig. 8f). In sinuous parts of the aggregates, the needles, however, have to be arranged with a slight twist to one another, so as to describe the given curvature. Finally, we note that on occasion the surface of the aggregates becomes overcast by an array of blocky rhombohedra (see Fig. 4d), most probably as a consequence of secondary crystallization.

To understand the role of the silica during the crystallization process, it is crucial to realize that not the whole amount of silica added is coprecipitated with

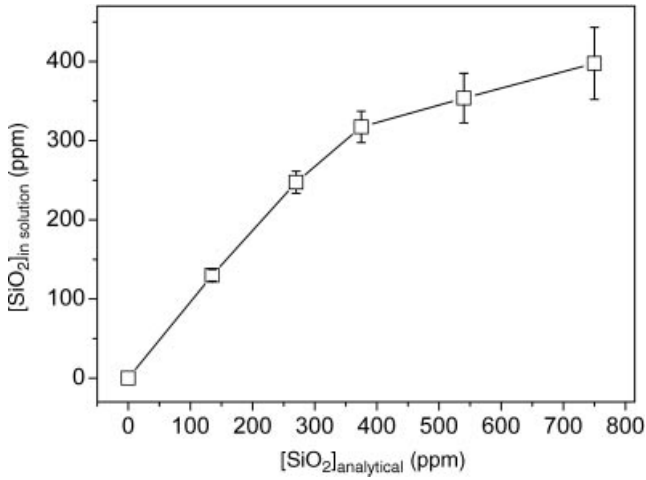


Figure 9. Estimated concentrations of dissolved silica in the mixtures after 5 min and corresponding standard deviations, outlined as a function of the respective total SiO_2 content. Values were obtained by calculating the amount of precipitated silica on the basis of the Si/Ca ratio measured by EDX spectroscopy, premising that the present 5 mM $CaCO_3$ were completely converted to ACC.

the ACC particles at the beginning of the experiments. From the Si/Ca atomic ratio determined by EDX for the amorphous material isolated 5 min after mixing, the concentration of silica remaining dissolved in solution at that time can be calculated when assuming, in an approximation, that Ca^{2+} reacts quantitatively to ACC. Results are shown in Fig. 9. At 135 and 270 ppm SiO_2, less than 10% of the available silicate species have in fact been deposited on the ACC nanoparticles, while around 15% have precipitated at 375 ppm. The fraction of silica removed from solution increases considerably when further raising the overall content to about 35% and 47% at 540 and 750 ppm, respectively. This implies in turn that at least more than half of the introduced molar amount of silica exists as soluble mono- and oligomers in the systems after initial ACC particle formation. Thus, data evidence that addition of more silica not only entails longer term stabilization of ACC precursors, but also implicates that crystallization of $CaCO_3$ occurs under the influence of a higher concentration of dissolved silicate ions in solution. On the other hand, once crystallization has commenced, ongoing redissolution of empty silica skins should cause a steady increase of the concentration in solution with time and as growth proceeds.

D. Crystal Polymorphism and Composition

Powder diffraction patterns of crystals recovered from samples at 135 and 270 ppm SiO_2 demonstrate that the only polymorph formed in the experiments was calcite,

as expected in view of the observed rhombohedral habit (cf. Fig. 6). SEM analyses further prove that the small fraction of vaterite particles traced in the silica-free reference was consistently absent in the presence of silica. This may indicate that added silica promotes the formation of calcite relative to vaterite under the given conditions and would hence support the results of a recent study on the effect of dissolved silica on $CaCO_3$ crystallization [71]. The polymorphism of structures generated at higher silica concentrations was investigated by monitoring the diffraction of selected individual particles. Diffraction images of the morphologies obtained at 375 ppm SiO_2 typically display multiple pairs of isolated spots along various virtual rings (Fig. 10a). Such behavior argues for the existence of several larger crystallographically coherent domains maintaining different orientations. This suggests that each of the projections eventually developing threefold symmetry is of single-crystalline nature (cf. Fig. 7c). As opposed to that, patterns obtained

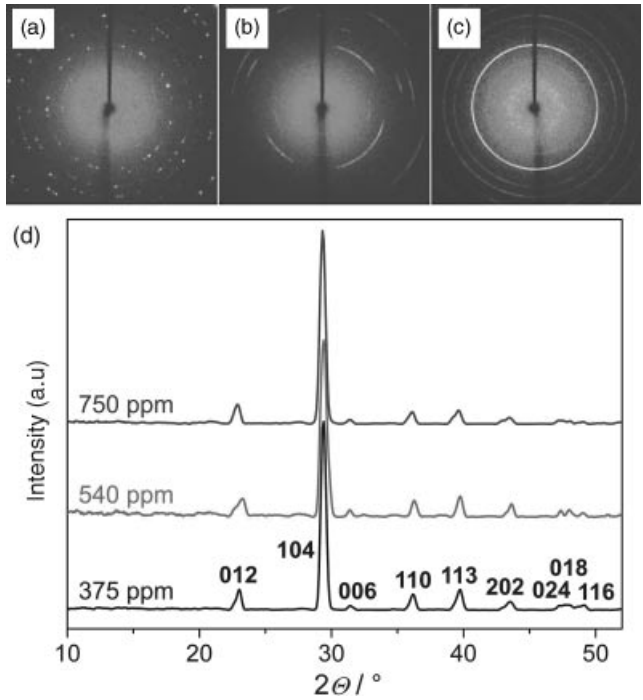


Figure 10. Single X-ray diffraction images acquired from (**a**) a three-pointed star-decorated spherulite formed at 375 ppm (Fig. 7c), (**b**) a dumbbell consisting of fibrous microneedles at 540 ppm (Fig. 8a), and (**c**) an open porous spheroidal structure at 750 ppm SiO_2 (Fig. 8b). (**d**) Diffractograms derived by summing up all diffraction images collected for a given specimen (after background subtraction) and integrating the intensity radially over all angles. The occurring reflections can be assigned to the calcite lattice.

for the architectures grown at 540 and 750 ppm SiO_2 show arc-like reflections (Fig. 10b) or full rings (Fig. 10c) and thus confirm that these samples are polycrystalline aggregates. Streaking of spots was thereby observed only for dumbbell-like specimens and can hence likely be ascribed to the shape anisotropy of this kind of morphology—in contrast to the quasi-isotropic geometry of spherulitic aggregates that gives continuous rings. These differences hint at distinct morphogenetic pathways responsible for the formation of extraordinary structures retrieved at 375 and 540/750 ppm silica, respectively, although both share a phenomenological growth sequence involving elongated precursors, dumbbell-shaped intermediates, and spherulitic terminal morphologies. Furthermore, despite the fact that the rod-like habit of the crystallites constituting the aggregates isolated at higher silica concentrations is commonly associated with the aragonite modification, analyses of the diffraction data evidence that all products were pure calcite and no other phase was present (Fig. 10d).

The degree of silica incorporation to the emerging crystal architectures was studied by measuring the content of Si relative to Ca for a series of representative structures at each concentration by means of EDX spectroscopy. For this purpose, specimens were washed carefully prior to measurement in order to remove potentially adhering silica-containing amorphous floccules. Results were found to be well reproducible. Averaged values for the Si/Ca atomic ratio are plotted in Fig. 11. The fraction of silica included in the final crystals is generally rather small

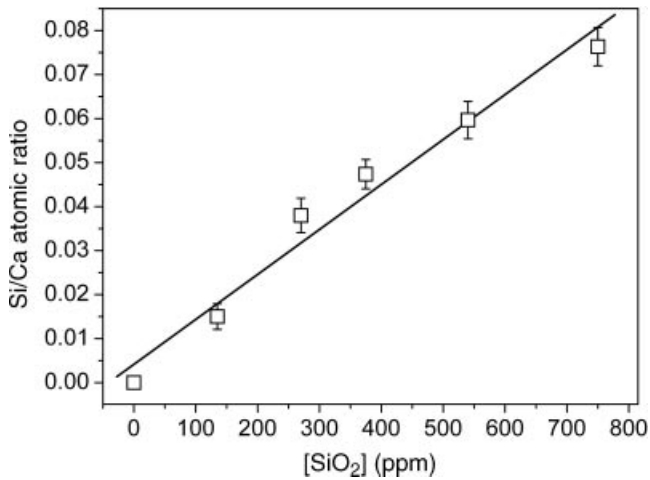


Figure 11. Mean silicon content of characteristic crystalline morphologies obtained from samples at different silica concentrations, given by the Si/Ca atomic ratio determined by EDX and its standard deviation. The full line is a linear fit to the data ($Si/Ca = (4 \pm 4) \times 10^{-3} + (1.02 \pm 0.09) \times 10^{-4} \cdot [SiO_2]/ppm$).

(Si/Ca < 0.1) compared to the ACC precursor particles (cf. Fig. 2d), but apparently scales in more or less linear manner with the analytical amount added, though a slight trend toward saturation can be discerned at higher concentrations. This implies that reinforced interactions between the growing carbonate crystals and dissolved silicate species are to be expected when increasing the total content of silica in the samples.

IV. DISCUSSION

The data collected in this work suggest a scenario as illustrated in Fig. 12. Upon combining 10 mM solutions of calcium chloride and sodium carbonate, the created high supersaturation leads to rapid nucleation of calcium carbonate and the formation of spherical ACC nanoparticles throughout the mixture. When no silica is present, the lifetime of these precursor species is quite short and stable calcite

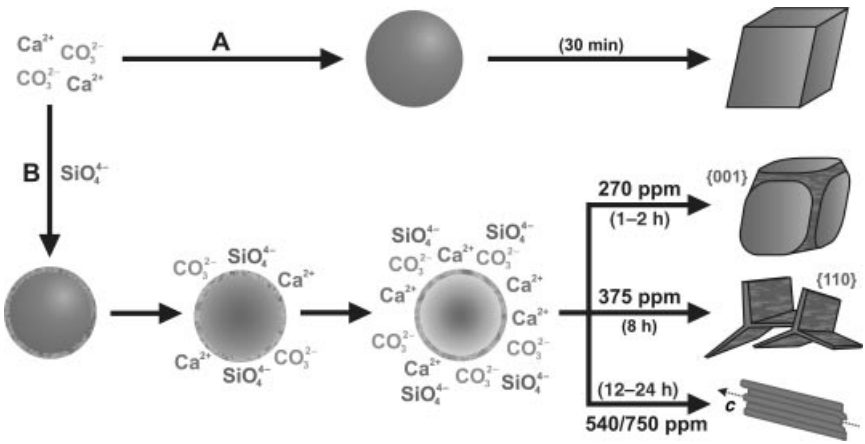


Figure 12. Crystallization of calcite under precipitation conditions in the absence (path A) and presence of different amounts of sodium silicate (path B). Added silicate precipitates spontaneously on growing ACC particles (spheres) forming continuous layers of amorphous silica all over their surface (indicated as marbled shell). These shells impede exchange with the surrounding medium, thus decelerating ACC redissolution and transformation to calcite. Crystallization is protracted the longer, the higher the silica concentration, such that more time is granted for structure evolution. By this means, dissolved silicate interferes in distinct ways and to varying degrees with crystal growth, resulting in morphologies of increasing complexity (end of path B, hachured faces are preferential sites for silica adsorption and precipitation). Times in brackets specify the approximate period after which first mature crystals or crystal aggregates were observed in the respective sample. For the sake of simplicity, the term "SiO_4^{4-}" is meant to represent all soluble silicate species present in the systems and bicarbonate ions existing in equilibrium with carbonate are not depicted in the sketch. Note further that structures are not drawn to relative scale.

crystals exhibiting classical rhombohedral equilibrium morphologies develop at their expense after a delay on the order of minutes (path A in Fig. 12). On addition of silica, protons generated at the surface of growing ACC particles due to dissociation of bicarbonate ions virtually decrease the local pH relative to the alkaline bulk medium ($pH_{bulk} \approx 10.6$) and thus trigger polymerization of silica in the close vicinity [67]. This results in the deposition of silica around the nanoparticles (path B in Fig. 12), the thickness of which increases with the amount of silica added. In the studied range of silica concentrations, these outer layers are porous and hence pervious to reagents (except for the fraction of long-term stable core–shell particles at 750 ppm SiO_2). Nevertheless, coating of the ACC precursors in silica clearly restricts the accessibility to the solvent such that dissolution of ACC and concurrent growth of crystalline phases are slowed down. Therefore, $CaCO_3$ units are released only gradually from the core volume. Since the overall porosity of the skins tends to be lower at higher silica content [67], the rate of crystallization becomes reduced incrementally as more silica is added and its onset is consistently retarded.

Based thereon, excess silica in solution is able to influence the growth behavior of calcite crystals and induce extraordinary morphologies and textures, which are discussed in the following. Generally, the precipitation conditions chosen in this study are not favorable for the design of complex crystal architectures, since additives have barely time to intervene with the growth process. However, with the embedding of ACC particles in siliceous shells, the supersaturation of the system with respect to calcite becomes relieved. This temporary storage of $CaCO_3$ as protected solid ACC establishes a situation in which structured mineralization over time frames of hours or days is possible.

At 135 ppm SiO_2, there were no significant morphological changes discernible on the final crystals in comparison to the control experiment at 0 ppm, apart from irregular silica deposits on the surface of as-grown rhombohedra (Fig. 6a). This can presumably be ascribed to the relatively low concentration of dissolved silicate species in the system (cf. Fig. 9) or insufficient deceleration of crystallization, both preventing pronounced interactions between the developing crystals and the additive. Marked effects can in turn be distinguished starting from 270 ppm SiO_2. First, truncation of the sharp edges and corners usually delimiting the rhombohedral shape is observed, causing a rounded appearance and indicating the expression of uncommon $\{001\}$ faces (Fig. 6b). Similar structures have been reported previously for calcite crystals grown in the presence of polystyrene sulfonate [72] or double-hydrophilic copolymers comprising an anionic block of aromatic sulfonate groups [73]. The exposure of highly charged $\{001\}$ planes was thereby attributed to stabilization via selective adsorption of the polyelectrolyte additives. We suspect that higher silicate oligomers, generated during local polymerization and later in the course of shell redissolution, act as polyanionic impurities modulating crystal growth in an analogical manner and hence effectuating the found distortion of the rhombohedral habit.

By contrast, crystals forming in samples at 375 ppm SiO_2 experience elongation along their *c*-axes that occurs through successive splitting of an initial rhomb and the emergence of fibrous projections, which later develop into single-crystalline strands having three wings at mutual angles of 120° (Fig. 7). Calcite morphologies with such 3*m* rotational symmetry were first grown by García-Ruiz and Domínguez-Bella in alkaline silica gels [74]. Under similar conditions, Imai et al. obtained hierarchical calcite superstructures consisting of three-pointed star-like subunits [75]. Both elongation and the subsequent formation of projections with threefold symmetry were explained on the basis of preferential adsorption and condensation of silicate ions on the {110} planes parallel to the *c*-axis (cf. Fig. 12). A gradually growing influence of the silica was thereby proposed to induce the transition from fibers to three-pointed stars. Most likely, a similar mechanism accounts for morphogenesis in the present case, especially in light of the finding that calcite crystallization from ACC precursors is accompanied by redissolution of silica skins and hence a continuous increase in the concentration of soluble silicate ions during growth.

As opposed to these elaborately shaped single crystals, growth of calcite in the presence of 540 and 750 ppm SiO_2 affords polycrystalline aggregates comprising a multitude of co-oriented microneedles (Fig. 8). Though not common, needle-like calcite crystals have been reported in the literature. For instance, acicular morphologies are well known in pedology, given that the so-called "needle-fiber calcite" (NFC) represents one of the major fractions of calcium carbonate found in vadose soil environments [76–78]. The formation of elongated crystals is thereby thought either to be caused by drastic physicochemical conditions during crystallization (such as fast evaporation) or to originate from the influence of biogenic matter (e.g., mineralization in plant tissues or bacterial activity). Another example for naturally occurring needle-shaped calcite is the teeth of sea urchins where micron-sized rods grow from a base plate along the unusual [102] direction [79]. Synthetically, spicular calcite crystals have been prepared via a liquid-precursor pathway [80], through spherulitic growth from a central core of polysaccharide-stabilized ACC [81], or by addition of certain acrylate/styrene copolymers [82]. In the latter case, morphological control was proposed to be achieved by specific adsorption of the polymeric additives on a subset of planes parallel to the *c*-axis, onto which the carboxylate chains match stereochemically. Condensed silicate oligomers probably perform related functions and block crystal faces parallel to *c*, thus causing the observed elongation of the crystallites.

Apparently, there is a change in the mode of interaction between growing calcite crystals and added silica when increasing the SiO_2 concentration. At lower silica content (270 ppm), it is the highly charged {001} faces perpendicular to the *c*-axis that become subject to pronounced adsorption, whereas at higher concentrations (375–750 ppm) planes parallel to *c* such as {110} are stabilized. Adsorption on {001} is generally considered to be dominated by nonspecific electrostatic binding,

while deposition of additive molecules on {110} faces typically requires a geometrical conformance of the molecular structure and the arrangement of ions in the respective crystal plane (i.e., the position and orientation of carbonate groups). This indicates that silica exists in different speciations in the samples despite their virtually identical bulk pH, possibly as a consequence of distinct degrees of condensation resulting from local polymerization and dissolution processes or due to the varying concentration itself.

When comparing the structures formed at 375 and 540/750 ppm, there are further peculiarities worth noting. At the lower concentration, the silica initially provokes splitting of a rhombohedral crystal that is, most probably, induced by insular blocking of the {001} faces. In the following, fibrous projections grow from the parent crystal maintaining its crystallographic orientation, with silicate species adsorbing strongly on multiple planes parallel to *c*. Eventually, adsorption on {110} becomes particularly favored and three-pointed stars are generated. Thus, by stabilizing distinct faces to a different extent depending on the stage of growth, the silica accretively deforms the regular calcite habit producing fractal architectures of predominantly single-crystalline nature. At higher silica concentrations, elongation along the *c*-axis occurs straight after nucleation as the effects discussed above are significantly more pronounced. In addition, crystal growth is effectively abandoned once the needles have reached lengths in the range of some microns, yielding fairly monodisperse crystallites that assemble under the conditions of the experiments to build up intriguing shapes as shown in Fig. 8. Comparable uniform nanocrystals are known to serve as building blocks constituting the various crystal architectures displayed by silica biomorphs [55, 58–60, 64–66]. In this case, the stability of the crystallites against ripening is believed to be facilitated by local silica polymerization and consequent coating of the particles [65, 66]. We speculate that such a scenario also accounts for the stabilization of calcite microneedles in the present systems; that is, specific adsorption of silicate oligomers on certain crystal faces passes into extensive silica precipitation over the entire surface at some point. This notion is supported by the measured relative content of Si in the aggregates (cf. Fig. 11), which complies with values reported for silica biomorphs [83, 84]. The resulting silica deposits prevent crystallites from growing larger, which is in full analogy to the stabilization of ACC particles traced immediately after the onset of $CaCO_3$ precipitation, where the presumed silica skins could be experimentally verified (cf. Figs. 2a–d and 5).

The origin of the morphologies found at 540 and 750 ppm SiO_2 remains uncertain. The observed "rod-to-dumbbell-to-sphere" transition has been described in the literature for biomimetic fluorapatite-gelatin composites, where morphogenesis was discussed in the framework of potential intrinsic electric fields [85]. A similar morphological evolution was recently identified to characterize the early stages of the formation of silica biomorphs [65, 66]. In this case, a single-crystal core undergoes continuous branching at noncrystallographic angles due to poisoning by oligomeric silicate impurities, which ultimately gives fractal spherulitic clusters.

The aggregates prepared in the present study indeed exhibit related shapes, but are evidently composed of fibrous subunits from the very beginning (see Fig. 4c). Interestingly, there are further analogies between the polycrystalline architectures of the present work and silica biomorphs. First, the open porous structure of the spherulite shown in Fig. 8b is reminiscent of distinct flower- and coral-like morphologies occasionally displayed by $BaCO_3$ and $CaCO_3$ biomorphs [59–61]. Second, the eventual reduction of global growth dimensionality to one (i.e., the formation of linear outgrowths) is a common feature of the two systems (see Fig. 8c) [63], although the striking filaments of silica biomorphs are clearly distinguished by their regular winding and helicity [55, 57–59, 64–66]. Finally, the arrangement of the crystallites within the aggregates deserves attention. In biomorphs, the nanorods are largely co-oriented with respect to their long axis, but maintain a coherent mutual misalignment particularly in curved segments. This leads to specific orientational ordering of the building units over extended scales [59]. Fairly akin structural principles are encountered in the $CaCO_3$ precipitates isolated at high silica concentrations in the present experiments (cf. Fig. 8d–f). This suggests that particle co-orientation is governed by similar interaction forces, which is reasonable when assuming that in both cases the crystallites are covered by silica. Thus, for lack of an adequate model explaining morphogenesis, we note that the crystal aggregates obtained at 540/750 ppm SiO_2 resemble silica biomorphs to a certain extent in terms of morphology and texture, with the marked distinction that the former consist of calcite microneedles while the latter are constructed from nanorods of aragonite-type carbonates.

V. CONCLUSIONS

Taken together, our experiments have demonstrated that adjusting the concentration of added silica during precipitation of calcium carbonate from solution is a simple but effective way to tune the stability of the initially formed ACC phase and, at the same time, create a variety of unusual calcite morphologies. During the early stages of precipitation, dissolved silicate species respond to pH changes occurring in the vicinity of growing ACC particles due to re-equilibration of the carbonate speciation by local polymerization and the deposition of silica skins all over the surface of the particles, which impedes energetically favored transformation to crystalline polymorphs. As crystallization sets in after increasing periods of delay, excess silica in solution and oligomers redissolved upon disintegration of the core–shell particles progressively affect calcite crystal growth as a result of combined adsorption and precipitation phenomena. Preceding stabilization of ACC precursors and the associated gradual supply of the medium with $CaCO_3$ units are essential prerequisites for concerted structure evolution in this case, as fast precipitation at the given high supersaturation would otherwise prevent any significant influence of the silica. The use of stabilized ACC as temporary storage

depot for $CaCO_3$ to lower the concentration of ions (and hence osmotic stresses) in solution as well as their subsequent release on demand are well-established concepts suggested to be applied *in vivo* during biomineralization [1–4]. In this regard, though purely inorganic, the studied systems mimic biogenic ones not only with respect to the formed morphologies, as put forward already in previous work [62–64]. Rather, the present data illustrate for the first time that there are further fundamental analogies in terms of the mechanisms underlying nonclassical crystallization in both cases.

On that basis, direct mixing of the components afforded crystalline structures exhibiting a complexity previously observed only for products obtained from syntheses under controlled conditions [74, 75]. Specific adsorption and condensation of silicate species on distinct planes of the calcite lattice and the concurrent alteration of relative growth rates led to an accretive deformation of the regular rhombohedral habit toward single crystals with threefold symmetry when increasing the amount of silica added. At high silica concentrations, enhanced adsorption caused rapid elongation of nuclei that became stabilized at microscale dimensions, probably through cementation in silica. The resulting needle-like crystallites were found to self-assemble into polycrystalline aggregates exhibiting intriguing noncrystallographic morphologies. Hence, the most prominent effect of the silica on calcite crystal growth is to stepwise miniaturize structural building units and thus enable hierarchical design, which again imitates strategies apparently employed by biomineralization to create materials with higher order textures [86].

Our findings highlight the singularity of the interplay between crystallizing carbonates and added silica, which, despite the simplicity of the reagents involved, is capable of inducing elaborate mineralization processes traditionally thought to rely on the presence of more complex organic additives, and show that (stabilized) amorphous precursors are also important intermediate species in these valuable laboratory model systems.

Acknowledgments

The authors thank Dr. Duane Choquesillo-Lazarte (LEC Granada) for performing X-ray diffraction studies, Dr. Regina Klein (University of Regensburg) for carrying out turbidity measurements, and Dr. Reinhard Rachel (University of Regensburg) for help with the TEM experiments. We are further grateful to Benjamin Gossler, Martina Heider, and Werner Reichstein (University of Bayreuth) for providing access to their field emission scanning electron microscope and support during the analyses. Financial support by the Fonds der Chemischen Industrie (M.K.), the program "Juan de la Cierva" of the Spanish Science Ministry (MICINN) (E.M.-G.), and the MICINN projects MAT2006-11701 and "Factoria de Cristalización" (J.M.G.-R.) is greatly appreciated.

References

1. H. A. Lowenstam and S. Weiner, *On Biomineralization*, Oxford University Press, New York 1989.
2. L. Addadi and S. Weiner, *Angew. Chem., Int. Ed. Engl.*, **31**, 153 (1992).

3. L. Addadi, S. Raz and S. Weiner, *Adv. Mater.*, **15**, 959 (2003).
4. S. Mann, *Biomineralization: Principles and Concepts in Bioinorganic Materials Chemistry*, Oxford University Press, New York, 2001.
5. E. Haeckel, *Art Forms from the Ocean: The Radiolarian Atlas of 1862*, Prestel Publishing, Munich/London, 2005.
6. G. A. Ozin, *Acc. Chem. Res.*, **30**, 17 (1997).
7. M. Sumper and E. Brunner, *Adv. Funct. Mater.*, **16**, 17 (2006).
8. O. Yamaguchi, T. Kanazawa, and K. Shimizu, K., *J. Chem. Soc., Dalton Trans.*, **5**, 1005 (1982).
9. D. Pontoni, J. Bolze, N. Dingenouts, T. Narayanan, and M. Ballauff, *J. Phys. Chem. B*, **107**, 5123 (2003).
10. H. Pan, X. Y. Liu, R. Tang, and H. Y. Xu, *Chem. Commun.*, **46**, 7415 (2010).
11. M. G. Taylor, K. Simkiss, G. N. Greaves, M. Okazaki, and S. Mann, *Proc. R. Soc. London, Ser. B*, **252**, 75 (1993).
12. J. Aizenberg, G. Lambert, L. Addadi, and S. Weiner, *J. Am. Chem. Soc.*, **124**, 32 (2002).
13. J. Aizenberg, G. Lambert, L. Addadi, and S. Weiner, *Adv. Mater.*, **8**, 222 (1996).
14. A. Al-Sawalmih, C. Li, S. Siegel, P. Fratzl, and O. Paris, *Adv. Mater.*, **21**, 4011 (2009).
15. A. Becker, U. Bismayer, M. Epple, H. Fabritius, B. Hasse, J. Shi, and A. Ziegler, *J. Chem. Soc., Dalton Trans.*, 551 (2003).
16. S. Weiner, J. Mahamid, Y. Politi, Y. Ma, and L. Addadi, *Front. Mater. Sci. Chin.*, **3**, 104 (2009).
17. Y. Politi, R. A. Metzler, M. Abrecht, B. Gilbert, F. H. Wilt, I. Sagi, L. Addadi, S. Weiner, and P. U. P. A. Gilbert, *Proc. Natl. Acad. Sci. USA*, **105**, 17362 (2008).
18. Y. Politi, T. Arad, E. Klein, S. Weiner, and L. Addadi, *Science*, **306**, 1161 (2004).
19. C. E. Killian, R. A. Metzler, Y. U. T. Gong, I. C. Olson, J. Aizenberg, Y. Politi, F. H. Wilt, A. Scholl, A. Young, A. Doran, M. Kunz, N. Tamura, S. N. Coppersmith, and P. U. P. A. Gilbert, *J. Am. Chem. Soc.*, **131**, 18404 (2009).
20. B. Hasse, H. Ehrenberg, J. C. Marxen, W. Becker, and M. Epple, *Chem. Eur. J.*, **6**, 3679 (2000).
21. L. Gago-Duport, M. J. I. Briones, J. B. Rodríguez, and B. Covelo, *J. Struct. Biol.*, **162**, 422 (2008).
22. H. A. Lowenstam, *Science*, **227**, 51 (1985).
23. J. Mahamid, B. Aichmayer, E. Shimoni, R. Ziblat, C. Li, S. Siegel, O. Paris, P. Fratzl, S. Weiner, and L. Addadi, *Proc. Natl. Acad. Sci. USA*, **107**, 6316 (2010).
24. A. Beniash, R. A. Metzler, R. S. K. Lam, and P. U. P. A. Gilbert, *J. Struct. Biol.*, **166**, 133 (2009).
25. N. Gehrke, N. Nassif, N. Pinna, M. Antonietti, H. S. Gupta, and H. Cölfen, *Chem. Mater.*, **17**, 6514 (2005).
26. E. Loste, R. J. Park, J. Warren, and F. C. Meldrum, *Adv. Funct. Mater.*, **14**, 1211 (2004).
27. C. Li and L. Qi, *Angew. Chem., Int. Ed.*, **47**, 2388 (2008).
28. F. C. Meldrum and H. Cölfen, *Chem. Rev.*, **108**, 4332 (2008).
29. F. C. Meldrum, *Int. Mater. Rev.*, **48**, 187 (2003).
30. F. Lippmann, *Sedimentary Carbonate Minerals*, Springer, Berlin, 1973.
31. R. E. Zeebe, J. C. Zachos, K. Caldeira, and T. Tyrell, *Science*, **321**, 51 (2008).
32. W. Tegethoff, J. Rohleder, and E. Kroker, *Calcium Carbonate: From the Cretaceous Period to the 21st Century*, Birkhäuser, Basel, 2001.
33. J. MacAdam and S. A. Parsons, *Rev. Environ. Sci. Biotechnol.*, **3**, 159 (2004).
34. H. Cölfen and M. Antonietti, *Mesocrystals and Nonclassical Crystallization*, Wiley, Chichester, 2008.

35. D. Gebauer, A. Völkel, and H. Cölfen, *Science*, **322**, 1819 (2008).
36. H. Cölfen and S. Mann, *Angew. Chem., Int. Ed.*, **42**, 2350 (2003).
37. J. Nyvlt, *Cryst. Res. Technol.*, **30**, 443 (1995).
38. L. Brecevic and A. E. Nielsen, *J. Cryst. Growth*, **89**, 504 (1989).
39. J. R. Clarkson, T. J. Price, and C. J. Adams, *J. Chem. Soc., Faraday Trans.*, **88**, 243 (1992).
40. J. Rieger, T. Frechen, G. Cox, W. Heckmann, C. Schmidt, and J. Thieme, *Faraday Discuss.*, **136**, 265 (2007).
41. T. Ogino, T. Suzuki, and K. Sawada, *Geochim. Cosmochim. Acta*, **51**, 2757 (1987).
42. J. Rieger, *Tenside Surf. Det.*, **39**, 6 (2002).
43. E. Loste, R. M. Wilson, R. Seshadri, and F. C. Meldrum, *J. Cryst. Growth*, **254**, 206 (2003).
44. D. C. Steytler, B. H. Robinson, J. Eastoe, K. Ibel, J. C. Dore, and I. MacDonald, *Langmuir*, **9**, 903 (1993).
45. B. Guillemet, M. Faatz, F. Gröhn, G. Wegner, and Y. Gnanou, *Langmuir*, **22**, 1875 (2006).
46. J. J. J. M. Donners, B. R. Heywood, E. W. Meijer, R. J. M. Nolte, and N. A. J. M. Sommerdijk, *Chem. Eur. J.*, **8**, 2561 (2002).
47. A. W. Xu, Q. Yu, W. F. Dong, M. Antonietti, and H. Cölfen, *Adv. Mater.*, **17**, 2217 (2005).
48. H. Cölfen and M. Antonietti, *Langmuir*, **14**, 582 (1998).
49. F. Jiang, Y. Yang, L. Huang, X. Chen, and Z. Shao, *J. Appl. Polym. Sci.*, **114**, 3686 (2009).
50. H. Cölfen, *Top. Curr. Chem.*, **271**, 1 (2007).
51. M. Antonietti, M. Breulmann, C. G. Göltner, H. Cölfen, K. K. W. Wong, D. Walsh, and S. Mann, *Chem. Eur. J.*, **4**, 2493 (1998).
52. H. Cölfen and L. Qi, *Chem. Eur. J.*, **7**, 106 (2001).
53. M. Li and S. Mann, *Adv. Funct. Mater.*, **12**, 773 (2002).
54. M. Li, B. Lebeau, and S. Mann, *Adv. Mater.*, **15**, 2032 (2003).
55. J. M. García-Ruiz and J. L. Amorós, *J. Cryst. Growth*, **55**, 379 (1981).
56. S. Domínguez-Bella and J. M. García-Ruiz, *J. Mater. Sci.*, **22**, 3095 (1987).
57. J. M. García-Ruiz, *Geology*, **26**, 843 (1998).
58. T. Terada, S. Yamabi, and H. Imai, *J. Cryst. Growth*, **253**, 435 (2003).
59. S. T. Hyde, A. M. Carnerup, A. K. Larsson, A. G. Christy, and J. M. García-Ruiz, *Physica A*, **339**, 24 (2004).
60. H. Imai, T. Terada, T. Miura, and S. Yamabi, *J. Cryst. Growth*, **244**, 200 (2002).
61. A. E. Voinescu, M. Kellermeier, B. Bartel, A. M. Carnerup, A. K. Larsson, D. Touraud, W. Kunz, L. Kienle, A. Pfitzner, and S. T. Hyde, *Cryst. Growth Des.*, **8**, 1515 (2008).
62. J. M. García-Ruiz, *Origins Life Evol. Biosphere*, **24**, 251 (1993).
63. J. M. García-Ruiz, A. M. Carnerup, A. G. Christy, N. J. Welham, and S. T. Hyde, *Astrobiology*, **2**, 353 (2002).
64. J. M. García-Ruiz, S. T. Hyde, A. M. Carnerup, A. G. Christy, M. J. van Kranendonk, and N. J. Welham, *Science*, **302**, 1194 (2003).
65. J. M. García-Ruiz, E. Melero-García, and S. T. Hyde, *Science*, **323**, 362 (2009).
66. W. Kunz and M. Kellermeier, *Science*, **323**, 344 (2009).
67. M. Kellermeier, E. Melero-García, F. Glaab, R. Klein, M. Drechsler, R. Rachel, J. M. García-Ruiz, and W. Kunz, *J. Am. Chem. Soc.*, **132**, 17859 (2010).
68. S. Schrödle, R. Buchner, and W. Kunz, *Fluid Phase Equilib.*, **216**, 175 (2004).

69. A. Gal, S. Weiner, and L. Addadi, *J. Am. Chem. Soc.*, **132**, 13208 (2010).
70. S. E. Wolf, J. Leiterer, M. Kappl, F. Emmerling, and W. Tremel, *J. Am. Chem. Soc.*, **130**, 12342 (2008).
71. L. Z. Lakshtanov and S. L. S. Stipp, *Geochim. Cosmochim. Acta*, **74**, 2655 (2010).
72. T. Wang, H. Cölfen, and M. Antonietti, *Chem. Eur. J.*, **12**, 5722 (2006).
73. A. N. Kulak, P. Iddon, Y. Li, S. P. Armes, H. Cölfen, O. Paris, R. M. Wilson, and F. C. Meldrum, *J. Am. Chem. Soc.*, **129**, 3729 (2007).
74. S. Domínguez-Bella and J. M. García-Ruiz, *J. Cryst. Growth*, **79**, 236 (1986).
75. H. Imai, T. Terada, and S. Yamabi, *Chem. Commun.*, 484 (2003).
76. E. P. Verrecchia and K. E. Verrecchia, *J. Sediment. Res.*, **64**, 650 (1994).
77. G. Cailleau, E. P. Verrecchia, O. Braissant, and L. Emmanuel, *Sedimentology*, **56**, 1858 (2009).
78. B. Bajnoczi, *Chem. Erde*, **66**, 203 (2006).
79. A. Berman, J. Hanson, L. Leiserowitz, T. F. Koetzle, S. Weiner, and L. Addadi, *Science*, **259**, 776 (1996).
80. M. J. Olszta, S. Gajjeraman, M. Kaufman, and L. B. Gower, *Chem. Mater.*, **16**, 2355 (2004).
81. C. Zhong and C. C. Chu, *Cryst. Growth Des.*, **10**, 5043 (2010).
82. L. Moore, J. D. Hopwood, and R. J. Davey, *J. Cryst. Growth*, **261**, 93 (2004).
83. A. E. Voinescu, M. Kellermeier, A. M. Carnerup, A. K. Larsson, D. Touraud, S. T. Hyde, and W. Kunz, *J. Cryst. Growth*, **306**, 152 (2007).
84. M. Kellermeier, F. Glaab, A. M. Carnerup, M. Drechsler, B. Gossler, S. T. Hyde, and W. Kunz, *J. Cryst. Growth*, **311**, 2530 (2009).
85. S. Busch, H. Dolhaine, A. DuChesne, S. Heinz, O. Hochrein, F. Laeri, O. Podebrad, U. Vietze, T. Weiland, and R. Kniep, *Eur. J. Inorg. Chem.*, 1643 (1999).
86. Y. Oaki, A. Kotachi, T. Miura, and H. Imai, *Adv. Funct. Mater.*, **16**, 1633 (2006).

AUTHOR INDEX

Numbers in parentheses are reference numbers and indicate that the author's work is referred to although his name is not mentioned in the text. Numbers in italic show the page on which the complete references are listed.

Advances in Chemical Physics, Volume 151: Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanoscale Materials, First Edition. Edited by Gregoire Nicolis and Dominique Maes.

SUBJECT INDEX

Advances in Chemical Physics, Volume 151: Kinetics and Thermodynamics of Multistep Nucleation and Self-Assembly in Nanoscale Materials, First Edition. Edited by Gregoire Nicolis and Dominique Maes.