Liquid–solid transition in nuclei of protein crystals

Aleksey Lomakin, Neer Asherie, and George B. Benedek*

Department of Physics, Center for Materials Science and Engineering and Material Processing Center, Massachusetts Institute of Technology, Cambridge, MA 02139-4307

Contributed by George B. Benedek, July 1, 2003

It is generally assumed that crystallization begins with a small, crystalline nucleus. For proteins this paradigm may not be valid. Our numerical simulations show that under conditions typically used to produce protein crystals, small clusters of model proteins (particles with short-range, attractive interactions) cannot maintain a crystalline structure. Protein crystal nucleation is therefore an indirect, two-step process. A nucleus first forms and grows as a disordered, liquid-like aggregate. Once the aggregate grows beyond a critical size (about a few hundred particles) crystal nucleation becomes possible.

The inability to produce high-quality crystals is currently a major hurdle in determining the 3D x-ray structure of proteins (1). In addition, the crystallization of proteins is involved in diseases such as human genetic cataract (2). It is therefore important to improve the understanding of the thermodynamics and kinetics of protein crystallization. Some progress has been made through the use of simple models that capture the essential features of protein interactions (3, 4). In particular, a globular protein is well described as a hard sphere with short-range, attractive interactions (5). The phase diagram of such model particles (Fig. 1) is believed to be generic for globular proteins. There are two possible phase transitions: crystallization, in which liquid and solid coexist, and liquid–liquid phase separation (LLPS), in which dilute and dense liquid phases coexist. LLPS is analogous to the gas–liquid phase transition observed for simple fluids, such as argon. In proteins, however, LLPS is metastable with respect to crystallization, whereas for simple fluids a stable liquid phase exists. This difference is due to the relatively large size of proteins: if the range of interaction is less than approximately a quarter of the particle diameter, there is no stable liquid phase (6–11). Because LLPS is metastable, it is not usually observed in protein solutions. Nevertheless, it has been studied systematically for lysozyme (12–14) and several γ-crystallins (15, 16). The full phase diagram of these proteins resembles that shown in Fig. 1 (17, 18).

Proteins are crystallized from highly supersaturated but dilute solutions to promote the nucleation rate for crystals while minimizing aggregation (19). At high supersaturations, small liquid droplets, although metastable, can also form. Because both liquid and solid phases can nucleate, the question arises as to the structure of a nucleus en route to forming a crystal.

Methods

Numerical simulations offer a way to study the structure of such small clusters. In a system with a fixed number of particles N and volume V, once a cluster has nucleated it will grow until it reaches a size at which it is in equilibrium with the surrounding phase. By choosing an appropriate N and a small enough V, we can obtain stable clusters of any desired number of particles. Furthermore, if the surrounding phase is dilute, the stable cluster will be enclosed in a volume that is much bigger than its own size. Placing a cluster in a finite volume that is sufficiently small to provide stabilization, but large enough so that boundary effects are negligible, is a natural way to study the thermodynamic properties of small clusters. It is an alternative to the biased sampling (20) commonly used in simulations to increase the probability of observing a nucleation event.

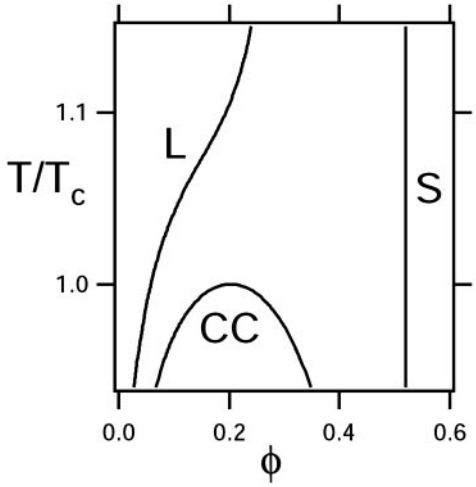

Fig. 1. The phase diagram of hard spheres with short-range attractive interactions. The reduced temperature T/T_c (T_c is the critical temperature) is shown as a function of the volume fraction ϕ. The liquidus (L) and solidus (S) lines mark the boundaries for the stable fluid and solid phases. The coexistence curve (CC) lies below the liquidus line, indicating that LLPS is metastable with respect to crystallization. The results in the figure are for an isotropic square-well potential (8, 22) with a reduced range $\lambda = 1.25$ (for which the reduced critical energy is $\varepsilon/kT_c = 1.27$), but very similar phase diagrams are obtained with other potentials if the range of interaction is sufficiently short (5–7).

The free energy F of N particles, of which M form a cluster whereas $N - M$ particles form the dilute phase (disconnected from the cluster) can be written as:

$$F(N, M) = F_c(M) + F_d(N - M). \quad [1]$$

Here F_c is the free energy of the cluster and F_d is the free energy of the dilute phase. Eq. **1** assumes that the interactions between the cluster and the dilute phase can be neglected.

The probability to observe a cluster of size M is given by (21)

$$P(M) = \frac{1}{Z(N)} \exp - \frac{F(N, M)}{kT}, \quad [2]$$

where $Z(N)$ is the partition function of the system, k is Boltzmann's constant and T is the absolute temperature. The chemical potential of a particle in the cluster $\mu_c(M) = dF_c(M)/dM$ is therefore

$$\mu_c(M) = -kT \frac{d \ln P(M)}{dM} + \mu_d(N - M). \quad [3]$$

In large systems, $P(M)$ is always close to its maximum value and $\mu_c \approx \mu_d$. However, in our case of small N and very small $N - M$, the large variation in $P(M)$ and μ_d because of the fluctuations in

Abbreviation: LLPS, liquid–liquid phase separation.

*To whom correspondence should be addressed at: Massachusetts Institute of Technology, Room 13-2005, 77 Massachusetts Avenue, Cambridge, MA 02139-4307. E-mail: gbb@mit.edu.

 www.pnas.org/cgi/doi/10.1073/pnas.1334069100

$N - M$ must be accounted for in the calculation of μ_c. This is accomplished by using the discretized version of Eq. **3** appropriate for small numbers of particles, i.e.,

$$\mu_c(M) = \mu_d(N - M) - kT \ln\left[\frac{P(M)}{P(M-1)}\right]. \quad [4]$$

Both the probability $P(M)$ and the chemical potential of a particle in the dilute phase μ_d can be calculated from the simulations: $P(M)$ from direct counting and μ_d by using Widom's method (22). In our case, the appropriate form of μ_d is

$$\mu_d(N - M) = \mu_0 - kT \ln\left[\frac{V}{\nu(N - M + 1)}\left\langle \exp\left(-\frac{\Delta E_{test}}{kT}\right)\right\rangle\right]. \quad [5]$$

Here μ_0 is the standard part of the chemical potential, ν is the hard core volume of a particle, ΔE_{test} is the change in the energy of the system attributable to the addition of a test particle and $\langle\ \rangle$ denotes a canonical ensemble average.

To examine the equilibrium properties of clusters of different sizes, we performed NVT (constant number of particles, volume, and temperature) Monte Carlo simulations of hard spheres with short-range interactions using an attractive square-well potential with reduced range λ and depth ε (22). For this potential, two particles are said to be in contact if their centers are within the range of interaction. A cluster is then defined as a group of particles forming a continuous series of contacts. By this definition, there is no energetic interaction between a cluster and the dilute phase. An entropic interaction, however, does exist, because clusters of different shape will exclude the particles in the dilute phase from slightly different volumes. If the cluster occupies only a small part of the simulation volume, which is the case when the difference between the densities of the cluster and the dilute phase is large, then the entropic interaction will be small. We find that Eq. **1**, which neglects all interactions between the cluster and the dilute phase, provides a good description of the systems we study.

We performed simulations for several short-range potentials (with $\lambda = 1.1$, 1.15, and 1.25), using a cubic volume with periodic boundary conditions. For each range, we examined a number of reduced energies ε/kT, chosen so that the system was in the LLPS region (22). Once a cluster had formed in a simulation, the cluster size was determined and the chemical potential and the average number of contacts per particle in the cluster were computed periodically. In addition to quantitative measurements, we monitored the structure of the cluster using stereoscopic images. (A simulation program that provides these images, together with the initial configurations of particles that illustrate the material in this article, is available from http://web.mit.edu/physics/benedek/clusters.html.)

Results and Discussion

The principal result of our simulations is that for small clusters only liquid-like structures are stable. For a crystalline cluster to form it must be large enough. This phenomenon is illustrated in Fig. 2. Fig. 2*a* shows a sequence of three cluster-size probability distributions, $P(M)$, obtained in three consecutive simulation runs. Initially, the clusters were disordered (circles); this distribution of clusters remained stable indefinitely. Four particles were then added to the system and another run was performed. Close to the end of this new run crystallization occurred, resulting in a shoulder in the distribution (triangles). The next run produced a distribution of crystalline clusters (squares); this distribution also remained stable indefinitely. Images of the clusters (Fig. 2*b*) before and after the transition illustrate that crystallization had taken place.

The transition between a liquid-like and a crystalline cluster results from the competition between the surface and the bulk

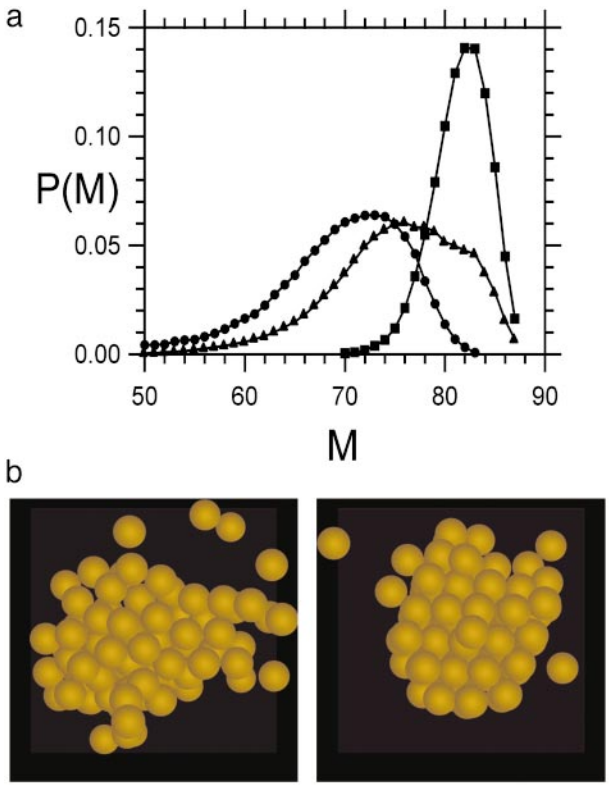

Fig. 2. Crystallization of a cluster with $\lambda = 1.1$ and $\varepsilon = 2.5kT$. (*a*) The probability distribution $P(M)$ to observe a cluster of size M at three different moments during the simulation. Our Monte Carlo simulations provide an artificial kinetic evolution of the system, allowing us to observe the transition between liquid and crystalline clusters. Initially, 84 particles are in the system and there is a stable distribution of sizes for the liquid cluster (●). Next, four particles are added to the system. After a while, crystallization occurs, leading to an increase in the size of the cluster (▲). Soon afterward, the system of 88 particles reaches a stable distribution for the crystalline cluster (■). (*b*) The liquid cluster at the beginning of the simulation with 84 particles (*Left*) and the crystalline cluster in the end of the simulation with 88 particles (*Right*). 3D versions of these images may be viewed either as a stereograph or using red/blue 3D glasses. (A simulation program that provides these images, together with the initial configurations of particles that illustrate the material in this article, is available from http://web.mit.edu/physics/benedek/clusters.html.)

of the clusters. Our simulations are performed under conditions for which LLPS is metastable; i.e., a bulk crystal is more stable than a bulk liquid phase. However, for clusters of finite size, surface effects become important, and these favor a liquid-like structure. As the cluster size increases, the relative number of particles on the surface decreases, the bulk properties can dominate, and the transition from a liquid cluster to a crystalline one occurs.

Although the most stable large clusters are crystalline, metastable liquid-like clusters may persist. In these cases it is possible to analyze quantitatively the competition between surface and bulk effects. Fig. 3 shows the results of many different simulations with $\lambda = 1.25$ and $\varepsilon = 2.0kT$. For sufficiently large clusters, two distinct values for the chemical potentials of clusters of the same size are observed. The metastable cluster with higher chemical potential has an unordered, liquid-like structure, whereas the stable state with lower chemical potential is crystalline. When the clusters are sufficiently small, only liquid-like clusters are observed. In large clusters both the chemical potential μ_c (Fig. 3*a*) and the number of contacts per particle n_c (Fig. 3*b*) are expected to behave asymptotically as $const_1 + const_2$

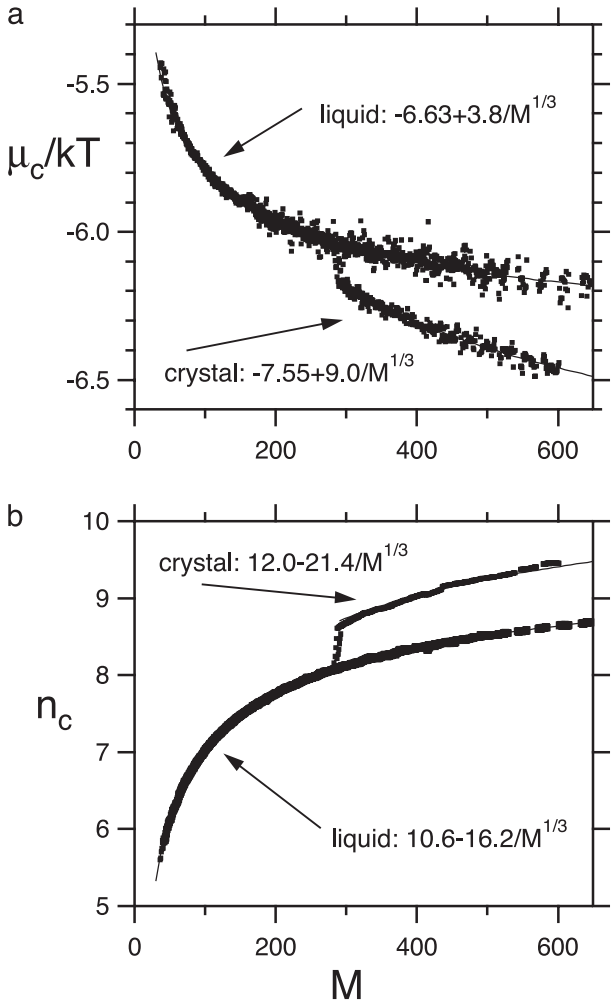

Fig. 3. Simulation results for $\lambda = 1.25$ and $\varepsilon = 2.0kT$ as a function of cluster size. The results are shown as black squares. (*a*) The chemical potential of a cluster. (*b*) The number of contacts per particle. Lines are fit to a functional dependence $const_1 + const_2 \times M^{-1/3}$ for the liquid and crystalline clusters.

$\times M^{-1/3}$, with the first term representing bulk properties and the second term representing surface properties (23). Lines in Fig. 3 are separate fits of the data to the formulae $\mu_c = \mu_{bulk} + \alpha_\mu M^{-1/3}$ and $n_c = n_{bulk} - \alpha_n M^{-1/3}$ for the crystalline and liquid clusters. Because the total energy of the cluster is $-Mn_c\varepsilon/2$, the energetic contribution to μ_{bulk} is $-n_{bulk}\varepsilon/2$ and to α_μ is $\alpha_n\varepsilon/3$.

The bulk chemical potential of crystalline clusters is lower than that of liquid-like clusters. This is because of a difference in n_{bulk}, i.e., an energetic difference; the entropies are essentially the same. In fact, the cell model (8), which assumes that each particle is caged in by its neighbors, reproduces fairly well the bulk chemical potential in both the crystalline and liquid clusters (Table 1).

Although the bulk free energy favors crystalline clusters, the surface free energy (proportional to α_μ) favors liquid clusters because these clusters have a much smaller surface free energy per particle. Therefore, sufficiently small clusters will be liquid-like, which may be viewed as a manifestation of surface melting (24, 25). As can be seen in Table 1, when the energy per contact is large, the smaller surface free energy in the liquid cluster arises from the large gain in entropy for a particle at the liquid surface as compared with the bulk. In contrast, there is practically no gain in entropy for a particle at the surface of a crystalline cluster. The surface free energy in the crystalline cluster is primarily due to the energetic contribution, i.e., loss of contacts at the surface.

Table 1. Bulk and surface characteristics of the clusters at $\lambda = 1.25$

ε/kT	n_{bulk}	μ_{bulk}/kT	α_μ/kT	$\alpha_n\varepsilon/3kT$	s/k	M_{trans}
	12	−9.30 (−9.04)	14.0	14.3	0.3	
2.2						225
	11	−7.74 (−7.94)	5.2	11.7	6.5	
	12	−7.55 (−7.84)	9.0	14.3	5.3	
2.0						280
	10.6	−6.63 (−6.44)	3.8	10.8	7.0	
	12	−6.10 (−6.64)	5.5	15.6	10.1	
1.8						440
	9.95	−5.62 (−4.80)	3.0	10.2	7.2	

The results in columns 2–6 are for both crystalline clusters (upper number) and liquid clusters (lower number). Column 1: reduced energy per contact. Column 2: the number of contacts in the bulk of the cluster. It is set to 12 for the crystalline clusters; for the liquid clusters the values are deduced from the fits of n_c. Column 3: bulk chemical potential of the cluster deduced from the fits of μ_c; in parentheses, the value predicted by the cell model (8), $\mu_{bulk} = -n_{bulk}\varepsilon/2 - 3kT\ln(\lambda - 1)$. Column 4: surface term in the chemical potential of the cluster. It is proportional to the free energy per surface particle. Column 5: the energetic contribution to α_μ/kT. Column 6: the entropic contribution to α_μ/kT. It is the difference between column 5 and column 4. Column 7: the typical size of the crystalline cluster at which the liquid-to-crystal transition is observed in the simulations. For a given range of interaction, this number decreases as the energy increases; at very large reduced energies, even small clusters are ordered.

Transitions between liquid-like and crystalline clusters have been demonstrated for 2D colloidal systems (26). In these experiments, spherical latex particles are trapped at an air–water interface. The addition of a single particle converts a colloidal cluster from a disordered structure to an ordered one. This type of transition is also well established for particles interacting through long-range potentials (24, 25, 27–30), which are used to describe clusters of rare gases, metals, and simple molecules. It is not surprising that liquid clusters form in long-ranged systems, where a stable liquid phase exists. It is, however, remarkable that liquid clusters also form in short-ranged systems, such as proteins, for which there is no stable liquid phase. Of course, because the interactions among real proteins are anisotropic, the stable structure of small clusters of proteins would most probably be an aggregate rather than a true liquid (31).

The results we have presented complement other studies of the possible role of metastable LLPS in the crystallization of proteins (32, 33). Those studies focused on crystal nucleation under conditions not usually used in protein crystallization experiments: high concentrations in the vicinity of the liquid–liquid coexistence curve, where the transition we describe does not exist.

Conclusions

Our findings suggest that a protein crystal should typically nucleate not as a small crystal but as a more stable disordered aggregate. Only when this aggregate reaches a critical size can it convert into a crystal. Aggregation *per se* is not detrimental to the outcome of a crystallization attempt. Instead, the crucial step is the ordering of a critical aggregate, which should occur faster than further unordered growth.

We thank O. Annunziata, S. Fraden, F. Lange, and G. Thurston for helpful discussions and critical comments. This work was supported in part by the National Institutes of Health (EY05127) and the National Aeronautics and Space Administration (NAG8-1659).

1. McPherson, A. (1999) *Crystallization of Biological Macromolecules* (Cold Spring Harbor Lab. Press, Plainview, NY).
2. Pande, A., Pande, J., Asherie, N., Lomakin, A., Ogun, O., King, J. & Benedek, G. B. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 6116–6120.
3. Piazza, R. (2000) *Curr. Opin. Colloid Interface Sci.* **5,** 38–43.
4. Anderson, V. J. & Lekkerkerker, H. N. W. (2002) *Nature* **416,** 811–815.
5. Rosenbaum, D., Zamora, P. C. & Zukoski, C. F. (1996) *Phys. Rev. Lett.* **76,** 150–153.
6. Hagen, M. H. J. & Frenkel, D. (1994) *J. Chem. Phys.* **101,** 4093–4096.
7. Lomba, E. & Almarza, N. G. (1994) *J. Chem. Phys.* **100,** 8367–8372.
8. Asherie, N., Lomakin, A. & Benedek, G. B. (1996) *Phys. Rev. Lett.* **77,** 4832–4835.
9. Rascón, C., Navascués, G. & Mederos, L. (1995) *Phys. Rev. B* **51,** 14899–14906.
10. Dannoun, A., Tejero, C. F. & Baus, M. (1994) *Phys. Rev. E* **50,** 2913–2924.
11. Illet, S. M., Orrock, A., Poon, W. C. K. & Pusey, P. N. (1995) *Phys. Rev. E* **51,** 1344–1352.
12. Ishimoto, C. & Tanaka, T. (1977) *Phys. Rev. Lett.* **39,** 474–477.
13. Broide, M. L., Tominic, T. M. & Saxowsky, M. D. (1996) *Phys. Rev. E* **53,** 6325–6335.
14. Galkin, O. & Vekilov, P. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 6277–6281.
15. Broide, M. L., Berland, C. R., Pande, J., Ogun, O. & Benedek, G. B. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 5660–5664.
16. Malfois, M., Bonneté, F., Belloni, L. & Tardieu, A. (1996) *J. Chem. Phys.* **105,** 3290–3300.
17. Berland, C. R., Thurston, G. M., Kondo, M., Boide, M. L., Pande, J., Ogun, O. & Benedek, G. B. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 1214–1218.
18. Muschol, M. & Rosenberger, F. (1997) *J. Chem. Phys.* **107,** 1953–1962.
19. Chernov, A. A. (1997) *Phys. Rep.* **288,** 61–74.
20. Frenkel, D. & Smit, D. (1996) *Understanding Molecular Simulation: From Algorithms to Applications* (Academic, San Diego).
21. Reif, F. (1965) *Fundamentals of Statistical and Thermal Physics* (McGraw–Hill, New York).
22. Lomakin, A., Asherie, N. & Benedek, G. B. (1996) *J. Chem. Phys.* **104,** 1646–1656.
23. Hirschfelder, J. O., Curtiss, C. F. & Bird, R. B. (1954) *Molecular Theory of Gases and Liquids* (Wiley, New York).
24. Briant, C. L. & Burton, J. J. (1975) *Nat. Phys. Sci.* **243,** 100–102.
25. Briant, C. L. & Burton, J. J. (1975) *J. Chem. Phys.* **63,** 2045–2058.
26. Onoda, G. Y. (1985) *Phys. Rev. Lett.* **55,** 226–229.
27. Lynden-Bell, R. M. & Wales, D. J. (1994) *J. Chem. Phys.* **101,** 1460–1476.
28. Berry, R. S. (1999) in *Theory of Atomic and Molecular Clusters*, ed. Jellinek, J. (Springer, New York), pp. 1–26.
29. Doye, J. P. K. & Calvo, F. (2001) *Phys. Rev. Lett.* **86,** 3570–3573.
30. Wales, D. J., Doye, J. P. K., Miller, M. A., Mortenson, P. N. & Walsh, T. R. (2000) *Adv. Chem. Phys.* **115,** 1–111.
31. Lomakin, A., Asherie, N. & Benedek, G. B. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9465–9468.
32. ten Wolde, P. R. & Frenkel, D. (1997) *Science* **277,** 1975–1978.
33. Costa, D., Ballone, P. & Caccamo, C. (2002) *J. Chem. Phys.* **116,** 3327–3338.