Aeolotopic interactions of globular proteins

ALEKSEY LOMAKIN, NEER ASHERIE, AND GEORGE B. BENEDEK*

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

ABSTRACT Protein crystallization, aggregation, liquid–liquid phase separation, and self-assembly are important in protein structure determination in the industrial processing of proteins and in the inhibition of protein condensation diseases. To fully describe such phase transformations in globular protein solutions, it is necessary to account for the strong spatial variation of the interactions on the protein surface. One difficulty is that each globular protein has its own unique surface, which is crucial for its biological function. However, the similarities amongst the macroscopic properties of different protein solutions suggest that there may exist a generic model that is capable of describing the nonuniform interactions between globular proteins. In this paper we present such a model, which includes the short-range interactions that vary from place to place on the surface of the protein. We show that this aeolotopic model [from the Greek *aiolos* ("variable") and *topos* ("place")] describes the phase diagram of globular proteins and provides insight into protein aggregation and crystallization.

Simple isotropic models that treat the protein molecules as spherical particles with short-range attractive interactions explain certain features of the protein phase diagram (1–5). In particular, liquid–liquid coexistence turns out to be metastable with respect to solidification when the range of interaction is less than one quarter of the particle diameter (6–10). This metastability has been observed for a variety of protein solutions (11–15) and in colloidal solutions (16–18), but not in simple fluids where the range of interaction is long (19). The isotropic model, however, fails to describe the phase diagram of protein solutions quantitatively and cannot address phenomena such as protein aggregation and self-assembly.

We use a simple model in which the energy of each particle depends only on its position relative to other particles and on its own orientation but is independent of the orientation of other particles. In this model, the pair potential of particles i and j has the form $w(\Omega_i, \Omega_j, \mathbf{r}_{ij}) = u(\Omega_i, \mathbf{r}_{ij}) + u(\Omega_j, -\mathbf{r}_{ij})$. Here, $\mathbf{r}_{ij}$ is the vector distance between particles i and j while Ω represents the three Euler angles that define the orientation of the particle. For such an additive model, it is possible to define the orientation-averaged free energy, $f_i(\{\mathbf{r}_{ij}\})$, of an individual particle as

$$\exp[-\beta f_i\{\mathbf{r}_{ij}\}] \equiv \int \exp\left[-\beta \sum_{j=1; j \neq i}^{N} u(\Omega_i, \mathbf{r}_{ij})\right] \frac{d\Omega_i}{8\pi_2}. \quad [1]$$

As usual, N is the number of particles, and $\beta = 1/k_B T$, where k_B is Boltzmann's constant and T is the absolute temperature. The free energy $f_i(\{\mathbf{r}_{ij}\})$ of a particle depends on the positions of all of the particles with which it interacts. If we can ignore correlations between these interactions, then $f_i(\{\mathbf{r}_{ij}\})$ can be written as $\Sigma_{j=1; j \neq i}^{N} U(r_{ij})$, where U is a temperature-dependent,

effective isotropic potential given by $\exp[-\beta U(r_{ij})] = \int \exp[-\beta u(\Omega_i, \mathbf{r}_{ij})] d\Omega_i / 8\pi^2$.

To study the conditions under which the aeolotopic potential $u(\Omega_i, \mathbf{r}_{ij})$ is "averageable," i.e., accurately approximated by the effective potential $U(r_{ij})$, we use the following modified square-well model. A protein molecule is represented by a spherical particle with a "map" of attractive regions covering a fractional area a of the surface. In this work, maps consisted of s non-overlapping spots of equal area on grids formed by equidistant meridional lines intersected by lines of constant latitude (see Fig. 1). As in the isotropic square well model (20), particles i and j are said to be neighbors if the distance r_{ij} between the centers of the two particles lies within $\sigma \leq r_{ij} < \lambda\sigma$ (where λ is the reduced range of interaction and σ is the diameter of the hard core). A particle i is said to make a contact with its neighbor j if the vector $\mathbf{r}_{ij}$ passes through an attractive spot on protein i. The energy $u(\Omega_i, \mathbf{r}_{ij})$ is defined as $-\varepsilon/2$ times the number of contacts made by the particle.

The effective pair potential for our model is a standard square-well potential with range λ and a temperature-dependent depth given by

$$\varepsilon_{eff}(T) = 2kT \ln[a\exp(\hat{\varepsilon}/2) + (1-a)], \quad [2]$$

where $\hat{\varepsilon} \equiv \varepsilon/kT$. To establish whether the aeolotopic potential is averageable, the reduced effective energy $\hat{\varepsilon}_{eff}(T) \equiv \varepsilon_{eff}(T)/kT$ is to be compared with $\hat{f} \equiv -2f_i(\{\mathbf{r}_{ij}\})/nkT$, the actual free energy per pair of neighbors (n is the number of neighbors).

We have numerically calculated $\hat{f}$ from Eq. **1** as a function of ε. Fig. 2 shows representative results for a particular map of 25 randomly placed spots with $a = 0.01$ and for two different numbers of neighbors: $n = 3$ and $n = 5$. For each n we show the results for five sets $\{\mathbf{r}_{ij}\}$ of randomly chosen positions of the neighbors. The black line is $\hat{\varepsilon}_{eff}(T)$.

For all configurations of the neighbors shown, the $n = 3$ results for $\hat{f}$ are close to $\varepsilon_{eff}(T)$, while for $n = 5$, the free energy $\hat{f}$ falls below $\hat{\varepsilon}_{eff}(T)$ at sufficiently large energies. This is because there are no orientations for which all five contacts are ever made. In fact, in four of the five configurations of neighbors, only three contacts were ever simultaneously made for this particular map. We see that, if the particle has three neighbors, its potential is averageable. For $n = 5$ it is not, and the full free energy must be explicitly calculated.

In Fig. 3a, we show the results of our simulations for the probability p_n that a particle cannot make all contacts with n randomly located neighbors for $n = 3$ (squares) and $n = 5$ (circles). Each point represents the results for a different map. As expected, maps with different patterns of the attractive spots, but the same s and a, have similar probabilities. The solid lines are given by

$$p_n(S, a) = \exp[-\kappa_n n^3 s^{3/2} a^{n-3/2}], \quad [3]$$

---

*To whom reprint requests should be addressed at: Room 13-2005, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139-4307. E-mail: gbb@mit.edu.
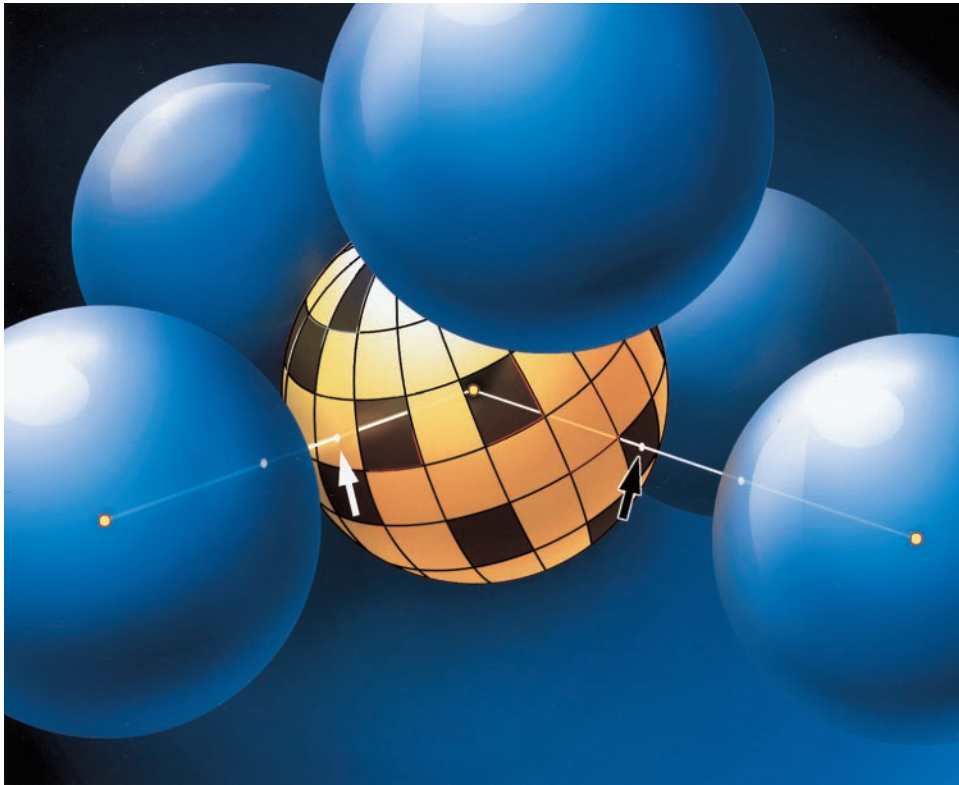
FIG. 1.   Schematic representation of the interactions of a model protein with its neighbors. The attractive spots on the central particle are shown as transparent. The central particle makes a contact with the front right neighbor but not with the front left neighbor.

where κ_n is of order one[†]. For the lines shown in Fig. 3*a*, we find $\kappa_3 = 0.67$ and $\kappa_5 = 0.80$.

When $p_n \approx 1$, only very specific configurations of neighbors allow contacts with all neighbors and the potential is not averageable, whereas when $p_n \approx 0$, all contacts are made in numerous orientations and the aeolotopic potential is averageable. In Fig. 3*b*, we present the simulation results for $p_n(s, a) = 0.5$ for $n = 3$ (squares) and $n = 5$ (circles). For maps whose s and a fall significantly above the open symbols (i.e., $p_n < 0.5$), the potential is averageable; below, it is not. The lines are obtained from Eq. **3** for $n = 3$ (solid) and $n = 5$ (dashed), assuming the values of κ_n given above. The solid triangle in Fig. 3*b* corresponds to the parameters of the map we have used in Fig. 2 ($s = 25$, $a = 0.01$).

The key parameter that determines whether an average potential can be used instead of an aeolotopic one is the number of neighbors n. Interactions that are averageable near the liquid–liquid critical point, where $n \approx 3$ (1), may no longer be averageable in the crystal, where $n > 5$. Thus, it is not always appropriate to use the same effective potential to describe both the liquid and solid phases.

This conclusion is borne out when we analyze the phase diagram of globular protein solutions. In Fig. 4, we show as open symbols the liquidus line (squares) and coexistence curve (circles) measured for the protein γ_{IIIb}-crystallin (11, 13). We have previously shown that the isotropic square well model with a reduced range of interaction $\lambda = 1.25$ produces the

correct coexistence curve (blue coexistence curve), provided that the depth of the potential has the appropriate temperature dependence (1). However, the liquidus line predicted by using the same potential in the solid (dashed red line) differs
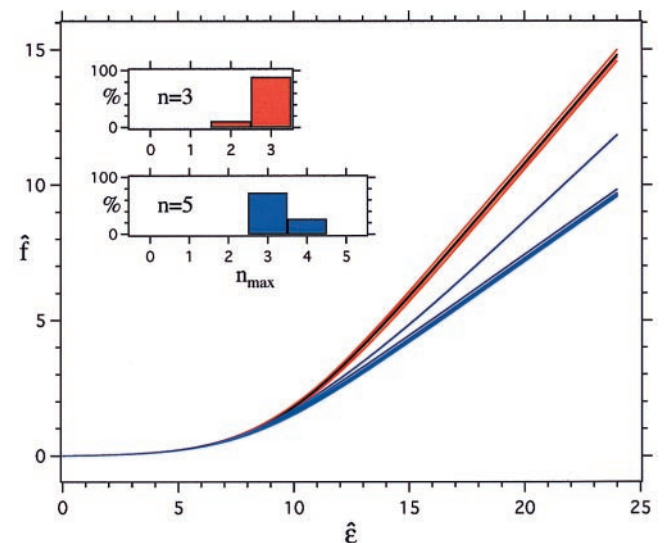

FIG. 2.   The reduced free energy as a function of the reduced energy. The black line is obtained assuming independent interactions. The red and blue lines are the simulation results for $\hat{f}$ with $n = 3$ and $n = 5$, respectively, for five different randomly chosen configurations of the neighbors. The same map, with $a = 0.01$ and $s = 25$, was used for both sets of simulations. For a given configuration of neighbors, the probability $p(m; n)$ that the central particle makes m contacts was calculated (for each m in the range $0 \leq m \leq n$) by directing counting of the contacts for 10 million random orientations of the central sphere. The reduced free energy then was obtained from $\hat{f} = (2/n)\ln[\Sigma_{m=0}^n p(m;n)\exp(m\hat{\varepsilon}/2)]$. The inset shows the percentage of simulations in which the maximum number of contacts made is n_{max}.
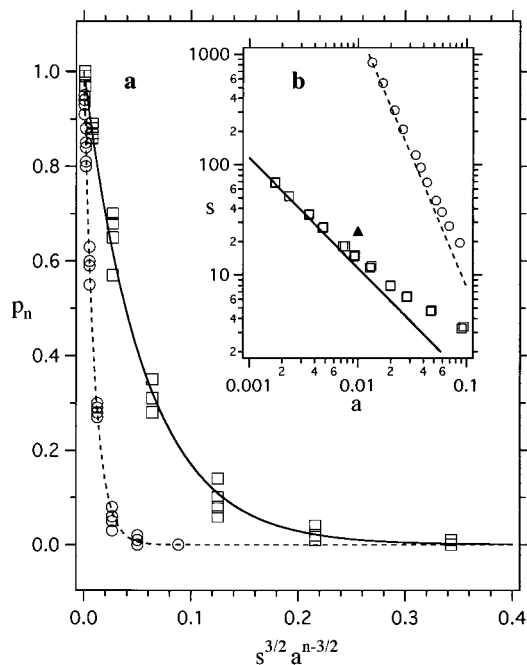
[†]One way to derive Eq. **3** is as follows. The probability that n neighbors make all n contacts for a particular orientation of the central sphere is a^n. This orientation allows a small rotational jiggle about each of the three axes of the sphere within an angle $\Delta\theta \sim (a/s)^{1/2}/n$ without losing any contacts. If two orientations differ by more than $\Delta\theta$, their probabilities of making all contacts are statistically independent (this last statement holds only for maps with many randomly placed spots). There are $1/(\Delta\theta)^3$ such independent orientations, and the probability that none of them makes all n contacts is $\exp(-a^n/(\Delta\theta)^3)$. Substituting for $\Delta\theta$, we recover Eq. **3**.

FIG. 3.   (*a*) The probability p_n of not making all *n* contacts. The open symbols are the simulation results for *n* = 3 (squares) and *n* = 5 (circles). Vertical clusters of points represent maps with the same *s* and *a* but different random patterns of the attractive spots. For a given configuration of neighbors, a large sampling of orientation space was performed to search for orientations of the central particle in which all *n* contacts were made. This search was performed by combining random reorientations of the particle with small rotations about a randomly chosen axis to ensure a thorough sampling of all possible orientations. The probability p_n was calculated by counting the number of configurations of neighbors (of 100 randomly chosen ones) in which the central sphere did not make all *n* contacts. The theoretical results for *n* = 3 (solid line) and *n* = 5 (dashed line) also are shown (see Eq. **3**). (*b*) The boundary for averaging the potential. The open symbols are the simulation results for *n* = 3 (squares) and *n* = 5 (circles). The values of *s* and *a* for each symbol are those that give $p_n = 0.5$. They were obtained by interpolating the simulation results for many different values of *s* and *a* that give $p_n \approx 0.5$. Above the open symbols, the potential is averageable for the corresponding *n*; below, it is not. The lines are the results expected for *n* = 3 (solid) and *n* = 5 (dashed) from Eq. **3**. The solid triangle corresponds to the values of *s* and *a* used in Fig. 2.

significantly from the experimental results (2). The aeolotopic model accounts in a natural way for the difference in apparent potentials in the two phases. The energy per particle that corresponds to the blue liquidus line (Fig. 4) is $-c\varepsilon/2 = -47k_BT_c$, where *c* is the number of contacts made per particle and T_c is the critical temperature (2, 21). For $c \approx 10$, we find that $\varepsilon \approx 9k_BT_c$. The effective energy at the critical temperature for $\lambda = 1.25$ is $\varepsilon_{eff}(T_c) = 1.3k_BT_c$ (1). Using these values of ε and $\varepsilon_{eff}(T_c)$, we find from Eq. **2** that $a \approx 0.01$. Remarkably, this value of *a* produces an $\varepsilon_{eff}(T)$ that yields the blue coexistence curve in Fig. 4. We see that the aeolotopic model provides a self-consistent and accurate representation of the liquidus line and the coexistence curve.

Aeolotopic interactions are responsible for protein aggregation. The number of neighbors within the aggregate is large, the energy per contact is much larger than the thermal energy, and the interactions are not averageable. All of these factors produce large energy barriers between the various conformations of aggregates which prevent the aggregates from rearranging into a crystal. Therefore, when the aeolotopic interactions are strong, metastable aggregates will abound, even though the crystal is the thermodynamically more stable
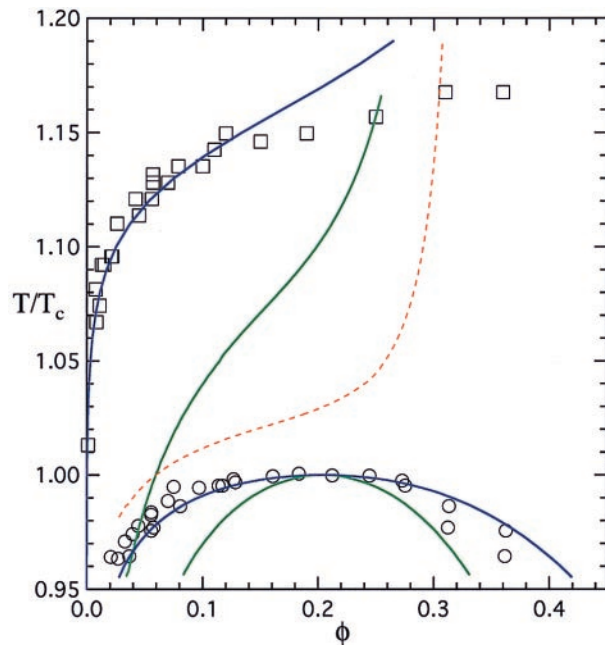

FIG. 4.   The phase diagram of γ_{IIIb}-crystallin. The open symbols are the experimentally determined liquidus line (squares) and coexistence curve (circles) for the protein. The green lines are the predictions for the boundaries assuming temperature-independent isotropic square well interactions (2). The blue coexistence curve is obtained for a square-well potential (1) with $\lambda = 1.25$ and $\varepsilon_{eff}(T)$ given by Eq. **2** with $\varepsilon = 9k_BT_c$ and $a = 0.01$. The blue liquidus line is obtained by equating the chemical potential of the liquid, with the above λ and $\varepsilon_{eff}(T)$, to the chemical potential of the solid with the parameters given in the text. The dashed red liquidus line corresponds to a solid interacting through the same effective isotropic potential $\varepsilon_{eff}(T)$ as the liquid.

structure. This view is consistent with the observation that, if the virial coefficient is too negative, amorphous aggregation dominates (22, 23). Of course, to properly address the competition between crystallization and aggregation, the kinetic properties of the aeolotopic model need to be investigated. However, the perspective we have presented suggests that it is the isotropic interactions between the proteins—both attractive and repulsive—which favor the formation of crystals over aggregates, while the aeolotopic interactions determine only the details of the crystal structure.

The aeolotopic model we have discussed captures the important features that characterize the interactions of proteins in solution. It provides a natural explanation as to why different apparent potentials are needed to describe the liquid and solid phases. Furthermore, the aeolotopic model lays the foundation for an analysis of phenomena unique to protein solutions, such as oligomerization, amorphous aggregation, and self-assembly.

1.   Lomakin, A., Asherie, N. & Benedek, G. B. (1996) *J. Chem. Phys.* **104,** 1646–1656.
2.   Asherie, N., Lomakin, A. & Benedek, G. B. (1996) *Phys. Rev. Lett.* **77,** 4832–4835.
3.   Rosenbaum, D., Zamora, P. C. & Zukoski, C. F. (1996) *Phys. Rev. Lett.* **76,** 150–153.
4.   Rosenbaum, D. F. & Zukoski C. F. (1996) *J. Crystallogr. Growth* **169,** 752–758.

5. Poon, W. C. K. (1997) *Phys. Rev. E*. **55,** 3762–3764.
6. Lekkerkerker, H. N. W., Poon, W. C. K., Pusey, P. N., Stroobants, A. & Warren, P. B. (1992) *Europhys. Lett.* **20,** 559–564.
7. Hagen, M. H. J., Meijer, E. J., Mooij, G. C. A. M., Frenkel, D. & Lekkerkerker, H. N. W. (1993) *Nature (London)* **365,** 425–426.
8. Hagen, M. H. J. & Frenkel, D. (1994) *J. Chem. Phys.* **101,** 4093–4097.
9. Daanoun, A., Tejero, C. F. & Baus. M. (1994) *Phys. Rev. E*. **50,** 2913–2924.
10. Rascòn, C., Navascués, G. & Mederos. L. (1995) *Phys. Rev. B* **51,** 14899–14906.
11. Berland, C. R., Thurston, G. M., Kondo, M., Broide, M. L., Pande, J., Ogun, O. O. & Benedek, G. B. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 1214–1218.
12. Thomson, J. A., Schurtenberger, P., Thurston, G. M. & Benedek, G. B. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7079–7083.
13. Broide, M. L., Berland, C. R., Pande, J., Ogun, O. O. & Benedek, G. B. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 5660–5664.
14. Broide, M. L., Tominc, T. M. & Saxowsky, M. D. (1996) *Phys. Rev. E*. **53,** 6325–6333.
15. Muschol, M. & Rosenberger, F. (1997) *J. Chem. Phys.* **107,** 1953–1962.
16. Verduin, H. & Dhont, J. K. G. (1995) *J. Colloid Interface Sci.* **172,** 425–437.
17. Ilett, S. M., Orrock, A., Poon, W. C. K. & Pusey, P. N. (1995) *Phys. Rev. E*. **51,** 1344–1352.
18. Gast, A. P. & Russel, W. B. (1998) *Phys. Today* **51,** 24–30.
19. Zemansky, M. W. (1968) *Heat and Thermodynamics* (McGraw–Hill, New York).
20. Reichl, L. E. (1980) *A Modern Course in Statistical Physics* (Univ. of Texas Press, Austin, TX).
21. Tanford, C. (1961) *Physical Chemistry of Macromolecules* (Wiley, New York).
22. George, A., Chiang, Y., Guo, B., Arabshahi, A., Cai, Z. & Wilson, W. W. (1997) *Methods Enzymol.* **276,** 100–110.
23. Neal, B. L., Asthagiri, D. & Lenhoff, A. M. (1998) *Biophys. J.* **75,** 2469–2477.