

Statistical mechanics of solvophobic aggregation: Additive and cooperative effects

Seishi Shimizu^{a)} and Hue Sun Chan^{b)}

Department of Biochemistry and Department of Medical Genetics and Microbiology, Faculty of Medicine, University of Toronto, Toronto, Ontario M5S 1A8, Canada

(Received 28 November 2000; accepted 25 May 2001)

Effects of possible non-pairwise-additive interactions on solvophobic aggregation are analyzed. A simple lattice model of binary solution with attractive solute-solute interactions is introduced to delineate the role of multiple-body effects in solute clustering and aggregation. Additive (noncooperative), cooperative, and anti-cooperative intersolute interactions are modeled by multiple-solute potentials that are respectively equal to, more favorable than, and less favorable than the sum of pairwise solute interactions. Under appropriate conditions, pairwise additive interactions and even interactions with significant anti-cooperativity can lead to aggregation and demixing. Cooperative interactions are not necessary for solute aggregation. Similarities and differences between solute aggregation and hydrophobic collapse of proteinlike heteropolymers are investigated. On average, heteropolymer collapse transitions as a function of solvophobic composition are significantly less sharp than the corresponding solute aggregation transitions. This difference is seen as a direct consequence of chain connectivity constraints. © 2001 American Institute of Physics. [DOI: 10.1063/1.1386420]

I. INTRODUCTION: DRIVING FORCES FOR SOLVATION VERSUS DESOLVATION

The clustering and aggregation of chemical groups in poor solvents, especially those of the hydrophobic variety,¹ is the mechanism underlying many biomolecular processes such as membrane formation, protein folding, ligand binding,²⁻⁴ and protein misfolding.⁵ While the precise physical origins of the effective interactions (potentials of mean force⁶) among solvophobic solutes may be complex, the basic statistical mechanics of aggregation is believed to be straightforward: Whether solutes are dispersed in solution or demix and form aggregates is the outcome of a competition between the configurational (mixing) entropy to be gained by dispersion and the energetic advantage to be achieved from favorable effective intersolute interactions upon aggregation. Therefore, aggregation is less likely in a more dilute solution. But given a strong effective attraction among solutes, aggregation is bound to occur when solute concentration is sufficiently high.⁷ This simple conceptual framework applies even to solutions of salts, for which an increasing salt concentration generally leads to an increased degree of ion pairing.^{8,9}

Hydrophobic compounds are expected to have tendencies to associate in water, but clustering of nonpolar solutes in water has been difficult to observe in computer simulations of low-concentration solutions, as is the case in an early model solution system with 4 solutes and 339 waters of Rapaport and Scheraga.¹⁰ Subsequent studies have confirmed that relatively high concentration of solutes is necessary to observe aggregation in simulations.¹⁰⁻²² The difficulty in ob-

serving association of small hydrophobic solutes in early simulations and experiments has led to a long-standing distinction between “pair” and “bulk” hydrophobic interactions.^{13,14,23} Pair¹³ or pairwise¹⁴ hydrophobic interaction refers to that between two or a small number of nonpolar molecules that are well exposed to water, whereas bulk hydrophobic interaction is that among “large clusters of nonpolar groups as may, for instance, be found in the interior of a protein molecule.”²³ Energetics of bulk hydrophobic interactions are often characterized by the transfer of small solutes from water to an amorphous liquid oil phase.^{1,13,24-26}

A number of researchers have examined the relationship between pair and bulk hydrophobic interactions.^{10-23,27} However, even the most basic question of whether there is a favorable driving force for two small nonpolar solutes to associate has been controversial. For instance, observing that the association among the nonpolar solutes in their computer simulations is less than that predicted by random mixing, Watanabe and Andersen¹¹ advanced the provocative idea of “hydrophobic repulsion,” stating that water molecules actually serve to keep nonpolar solutes apart.^{11,16} In a similar vein, Wood and Thompson¹³ interpreted experimental osmotic second virial coefficients to mean that pair hydrophobic interaction weakens nonpolar solutes’ mutual attraction relative to that in the gas phase. However, Blokzijl and Engberts¹⁴ have questioned whether the sampling in the solution simulations were sufficient. More fundamentally, Lüdemann *et al.*²² pointed out that the virial-coefficient argument for the unfavorability of hydrophobic association was probably flawed because it did not take into account the spatial dependence of the potential of mean force, which invariably exhibits a minimum favoring pairwise hydrophobic contacts between small nonpolar solutes.

^{a)}Electronic mail: shimizu@dirac.med.toronto.edu

^{b)}Electronic mail: chan@arrhenius.med.toronto.edu

Many-body effects in hydrophobic interactions can be complex.²⁸ The pair versus bulk question is essentially about whether hydrophobic interactions are pairwise additive or there are significant multiple-body cooperative or anti-cooperative effects. A pertinent observation is that although the favorability of pairwise hydrophobic interaction in dilute aqueous solution was called into question, clustering and aggregation have been observed in Wallqvist's^{18–20} simulation using a higher concentration of nonpolar solutes (18 Lennard-Jones solutes with 107 waters). This raises the possibility that hydrophobic aggregation may require cooperativity and multiple-body effects, since isolated solute pairs appear to prefer water-separated configurations.^{18–20} Based on an extensive set of careful simulations with different nonpolar solute concentrations and a novel method for counting hydrophobic contacts, Tsai *et al.*¹² have also observed solute clustering at high concentrations, and concluded that hydrophobic aggregation is “cooperative.” [The definition of cooperativity by Tsai *et al.* is different from the one used here (later in this work).] The idea that hydrophobic interaction at 25 °C and 1 atm is cooperative is not supported by potential of mean force simulations of Rank and Baker²¹ showing that the contact interaction free energies among multiple nonpolar solutes are slightly anti-cooperative (see their Figs. 4 and 6). A subsequent investigation of Czaplowski *et al.*²⁹ comparing three-methane and two-methane potentials of mean force indicates a small cooperative effect (see their Figs. 7 and 8). But our recent simulation indicates that anti-cooperativity prevails over an extended range of methane configurations under ambient conditions.³⁰

In view of these seemingly conflicting results and interpretations, it is useful to reexamine the idea that cooperativity is required for solvophobic aggregation, and delineate the general role of multiple-body, nonadditive interactions in clustering and aggregation. We provide a brief outline of the pertinent basic formulation in Sec. II. It should be noted in the present context that Wertheim's theory of solute associations³¹ has been applied to ion triplet formation.⁹ But this method is based on a pairwise additive assumption and therefore does not address the issue of interaction cooperativity being considered here. Integral equation approaches to solvation provide molecular details. As a result, their formulation for three-body effects are complex.³² Our aim here is not to address molecular details of hydrophobic interactions. (Our recent effort in that regard is reported elsewhere.²⁷) Rather, the main goal of the present work is to elucidate the basic principles governing interaction cooperativity and aggregation. Following a long tradition of lattice statistical mechanics models of solvation that have been proven useful,³³ we adopt a simple two-dimensional square-lattice model of binary solution for this effort (Sec. III).

As mentioned earlier, a prominent motivation for studying small-solute solvation and aggregation has been their implications on polymer collapse³⁴ and protein folding.¹² However, often only a heuristic discussion is provided for the relationship between aggregation of individual solutes and chain monomers constrained to be spatially close to one another to begin with.¹² Here we attempt to inject additional quantitative elements into such discussions. To this end, we

compare in Sec. IV the present lattice solution model with a closely related lattice chain model to clarify the similarities, or lack thereof, between solute aggregation and the collapse of heteropolymer chains in poor solvents. Finally, a brief summary of our findings is given in Sec. V.

II. ADDITIVE AND COOPERATIVE INTERACTIONS: DEFINITIONS

We first set up an analytical framework to facilitate subsequent developments. Using standard McMillan–Mayer theory,^{35–37} the osmotic pressure Π of a binary (two-component) solution may be written as the virial expansion

$$\frac{\Pi}{RT} = \rho + B_2^* \rho^2 + B_3^* \rho^3 + \dots, \quad (1)$$

where RT is gas constant times absolute temperature and ρ is the average number density of solute molecules (the other component is termed “solvent”). B_2^* , B_3^* , \dots are osmotic virial coefficients, which are determined by the potential of mean force among the solute molecules. In particular, the second and third virial coefficients are

$$\begin{aligned} B_2^* &= -\frac{1}{2} \int d\mathbf{r}_{12} (e^{-W^{(2)}(\mathbf{r}_{12})/RT} - 1), \\ B_3^* &= -\frac{1}{3} \int d\mathbf{r}_{12} \int d\mathbf{r}_{13} (e^{-W^{(3)}(\mathbf{r}_{12}, \mathbf{r}_{13})/RT} \\ &\quad - e^{-[W^{(2)}(\mathbf{r}_{12}) + W^{(2)}(\mathbf{r}_{13}) + W^{(2)}(\mathbf{r}_{23})]/RT}) \\ &\quad - \frac{1}{3} \int d\mathbf{r}_{12} \int d\mathbf{r}_{13} (e^{-W^{(2)}(\mathbf{r}_{12})/RT} - 1) \\ &\quad \times (e^{-W^{(2)}(\mathbf{r}_{13})/RT} - 1)(e^{-W^{(2)}(\mathbf{r}_{23})/RT} - 1), \end{aligned} \quad (2)$$

respectively, where $\mathbf{r}_{ij} \equiv \mathbf{r}_j - \mathbf{r}_i$, and \mathbf{r}_i is the position of the i th solute. $W^{(k)}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_k) = W^{(k)}(\mathbf{r}_{12}, \mathbf{r}_{13}, \dots, \mathbf{r}_{1k})$ is the potential of mean force among k solutes, and $W^{(k)}$ is assumed to depend only on the solutes' relative positions. The k -solute interaction is additive if

$$W^{(k)}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_k) = \sum_{i=1}^k \sum_{j=i+1}^k W^{(2)}(\mathbf{r}_{ij}). \quad (3)$$

The system is entirely additive if this holds for all $k > 2$. For attractive $W^{(2)}(\mathbf{r}_{ij}) < 0$, the k -solute interaction is cooperative if $W^{(k)} < \sum_{i < j}^k W^{(2)}(\mathbf{r}_{ij})$, whereas $W^{(k)} > \sum_{i < j}^k W^{(2)}(\mathbf{r}_{ij})$ means that the interaction is anti-cooperative (or, equivalently, negatively cooperative). It is possible for an interaction to be cooperative for certain solute configurations (sets of \mathbf{r}_{ij} 's) but anti-cooperative for other solute configurations.

In general, the homogeneous mixture (nonaggregated) phase is stable when $\partial\Pi/\partial\rho \geq 0$ (Ref. 38). Therefore, as a first approximation, the spinodal condition

$$\left. \frac{\partial\Pi}{\partial\rho} \right|_{\rho=\rho_a} = 1 + 2B_2^* \rho_a + 3B_3^* \rho_a^2 + \dots = 0 \quad (4)$$

may be used to estimate the critical (transition) concentration ρ_a at the onset of aggregation and phase separation. It is intuitively obvious from this formulation that cooperative

multiple-solute interactions can shift the equilibrium in favor of aggregation. For example, a cooperative three-body potential of mean force $W^{(3)}(\mathbf{r}_{12}, \mathbf{r}_{13}) < W^{(2)}(\mathbf{r}_{12}) + W^{(2)}(\mathbf{r}_{13}) + W^{(2)}(\mathbf{r}_{23})$ implies that B_3^* is smaller (more negative) than when the interactions are additive [Eq. (2)]; and Eq. (4) indicates that a shift of B_3^* in the negative direction is likely to shift ρ_a to a smaller value.

III. A LATTICE MODEL OF AGGREGATION IN A BINARY SOLUTION

Consider a lattice model of binary solution with n solutes and $N-n$ solvents. The solute and solvent are of the same size and each occupies a single position in a completely filled lattice of N sites (with no free volume); hence ρ is equal to solute mole fraction $X=n/N$ here. Excluded-volume-violating configurations are forbidden, i.e., each lattice site can only accommodate one solute or one solvent. We start with the pairwise additive (noncooperative) case by assigning a two-solute contact energy (potential of mean force), viz., $W^{(2)}(\mathbf{r}_{ij}) = \mathcal{E} \leq 0$ when \mathbf{r}_{ij} is one of the z vectors between two nearest-neighbor lattice sites, $W^{(2)} = +\infty$ when $\mathbf{r}_{ij} = 0$ (because of excluded volume), and $W^{(2)} = 0$ otherwise. z is the lattice coordination number; $z=4$ for the square lattice used here. Boltzmann averages are computed from the partition function

$$Q = \sum_h g_s(h) e^{-\mathcal{E}h/RT}, \quad (5)$$

where $h=0,1,2,\dots$ is the total number of solute-solute nearest-neighbor lattice contacts, and the density of states $g_s(h)$ is the number of solution configurations with h such contacts. Solute association is monitored by using Monte Carlo sampling (see later in this work) to estimate the average number of solute-solute contact $\langle h \rangle = \partial A / \partial \mathcal{E}$, where $A \equiv -RT \ln Q$. Physically, we expect that there are more configurations [larger $g_s(h)$] when the solutes are dispersed [smaller h] than when they are aggregated [i.e., large h 's tend to be associated with small g_s 's].

Additivity of solvent-mediated forces can have at least two closely related but not necessarily identical meanings: Multiple-solute potentials of mean force are well approximated by (i) the sum of pairwise potentials of mean force, as defined by Eq. (3), or (ii) a linear function of one of few conventional surface-area measures of the solutes' solvent accessibility.³⁹ For (ii), the deviation from additivity depends on the definition of the area measure.^{21,27} Because geometrically the total buried surface area (area made solvent inaccessible) of a multiple-solute system often (though not always) corresponds roughly to the sum of surface areas of one solute buried by another solute considered one pair at a time, a correlation between the conformity to the two definitions of additivity is expected. For simplicity, we do not distinguish them in this work. The two definitions are identical for the additive lattice model here provided that the "solvent accessible area" of a solute is identified with the number of sides (out of z) of its lattice site that are in contact with a solvent site.

We consider also a simple augmentation to this model to explore the ramifications of interaction cooperativity on aggregation: For every connected cluster of $k > 3$ solutes, we assign a

$$\text{cooperative energy} = \mathcal{E}_{\text{coop}}(k-2). \quad (6)$$

In other words, $W^{(k)}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_k) = \sum_{i < j}^k W^{(2)}(\mathbf{r}_{ij}) + \mathcal{E}_{\text{coop}}(k-2)$ if the k solutes in question form a connected cluster but is not a part of a bigger cluster, otherwise $W^{(k)}$ is given by a sum of pairwise terms plus smaller cluster terms that are subject to the same general rule Eq. (6). As a first test of principles, we make the simplifying modeling assumption that this cooperative energy depends only on the size of the cluster but not its geometric shape. The model is cooperative when $\mathcal{E}_{\text{coop}} < 0$, anti-cooperative when $\mathcal{E}_{\text{coop}} > 0$, and additive when $\mathcal{E}_{\text{coop}} = 0$.

The model binary solution used here is an $N=40 \times 40$ two-dimensional square lattice with periodic boundary conditions. In general, phase behavior depends upon system size.⁴⁰ But we expect our simulations to capture the basic properties of the present simple model system. Standard Metropolis Monte Carlo techniques are used to sample solution configurations at different values of \mathcal{E} , $\mathcal{E}_{\text{coop}}$, and solute mole fraction n/N . At each Monte Carlo step, two solutes are randomly chosen and an exchange of their positions is accepted or rejected according to the Metropolis criterion.⁴¹ Solute positions are initialized randomly. In each run, the first 6.0×10^6 to 1.0×10^8 Monte Carlo steps are discarded to allow for proper equilibration. The subsequent 1.0×10^8 to 3.0×10^8 steps are then used for averaging. Simulations are performed at several temperatures. The thermodynamic properties of the model change significantly with temperature, especially with respect to demixing. Longer equilibration runs are required for systems that exhibit demixing. Aggregation or demixing is monitored by

$$\alpha_s \equiv \frac{\langle h \rangle}{n}, \quad (7)$$

where $\langle \dots \rangle$ denotes Boltzmann averaging. Clearly, $0 \leq \alpha_s \leq z/2$. The order parameter α_s characterizes the association among solutes; $\alpha_s \approx z/2$ signals aggregation and demixing.

Is cooperative multiple-body interactions necessary for collapse and clustering? In the protein folding context, the question as to whether hydrophobic interaction is additive or cooperative is of particular interest because it has recently been shown that additive hydrophobic-like interaction energies are insufficient to account for the folding transitions' high degrees of *thermodynamic* cooperativity as determined by calorimetry.^{42,43} Figure 1 shows that when the solute-solute "sticking" $-\mathcal{E}/RT$ is sufficiently strong, aggregation occurs in the absence of cooperative interactions. In those cases ($-\mathcal{E}/RT > 2.5$), a sharp transition marks the commencement of aggregation when solute mole fraction X is increased beyond a certain critical value. Figure 2 shows typical configurations of the "dispersed," "transition," and "collapsed" (aggregated) solution states that are reminiscent of the very similar pictures reported by Tsai *et al.*¹² for mo-

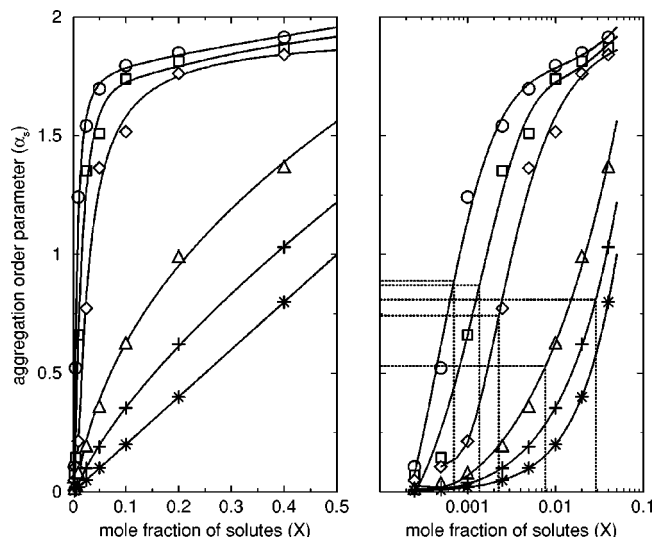


FIG. 1. Solvophobic solutes with strong pairwise additive ($\mathcal{E}_{\text{coop}}=0$) solute-solute interactions aggregate at high concentrations. α_s is plotted versus solute mole fraction X on a linear (left) and a logarithmic (right) scale. The latter provides a better resolution for the sharp transitions in α_s at small X . Data points are from Monte Carlo simulations, with \mathcal{E}/RT equals to (from top to bottom) -3.6 (circles), -3.0 (squares), -2.571 (diamonds), -1.5 (triangles), -0.75 (crosses), and 0 (asterisks). The solid line and curve through the asterisks are the relation $\alpha_s = zX/2$ predicted by random mixing without accounting for excluded volume. Monte Carlo simulations (asterisks) show that solute-solute contact probabilities in the $\mathcal{E}=0$ athermal case with excluded volume are also well described by this expression. The other fitted solid lines and curves are mere guides for the eye. The dotted lines in the plot on the right mark the critical densities $X = \rho_a = -1/2B_2^*$ and approximate critical α_s for solute aggregation predicted by the second osmotic virial coefficients.

lecular simulation of hydrophobic aggregation. Our present results show that cooperative intersolute interactions are not necessary for aggregation.

Results from lattice solution models may be compared to the McMillan–Mayer theory previously described. For the lattice model in this work, $B_2^* = [5 - 4 \exp(-\mathcal{E}/RT)]/2$ and $B_3^* = -12 \exp(-2\mathcal{E}/RT) [\exp(-\mathcal{E}_{\text{coop}}/RT) - 1] + 4 [\exp(-\mathcal{E}/RT) - 1]^2 + \frac{1}{3}$. Using Eq. (4), if only the first and the second osmotic virial terms are included in the expansion, $\rho_a = -1/2B_2^*$ expresses the critical solute density for aggregation. Figure 1 shows that this gives a reasonable description of the crossover from a dispersed to an aggregated regime: at $X = -1/2B_2^*$, the aggregation order parameters are predicted to be ≈ 0.53 – 0.89 (out of a range of 2) for the cases shown, and $X = -1/2B_2^*$ more or less marks the midpoints of the sharp transitions in α_s (as a function of X) for systems with strong solute-solute attractions ($\mathcal{E}/RT < 2.5$). However, because of its perturbative nature, the McMillan–Mayer expansion has limitations. For instance, no physical aggregation critical density ρ_a can be found for the additive systems in Fig. 2 if the third osmotic virial coefficient term is also included in Eq. (4) [in that case $\rho_a = (B_2^*/3B_3^*) (\sqrt{1 - 3B_3^*/(B_2^*)^2} - 1)$], contrary to the fact that aggregation does occur in these models. This is because for $\mathcal{E}_{\text{coop}}=0$ the above lattice expressions for B_2^* and B_3^* imply that $3B_3^* > (B_2^*)^2$.

Figure 3 shows the expected results that cooperative

inter-solute interactions ($\mathcal{E}_{\text{coop}} < 0$) facilitate aggregation in that they lower the critical mole fraction relative to the $\mathcal{E}_{\text{coop}}=0$ case, and that strong anti-cooperative intersolute interactions prevent aggregation. The more interesting observation, however, is that aggregation can occur even in the presence of a significant anti-cooperative term ($\mathcal{E}_{\text{coop}}/RT = +2.4$) when solute mole fraction is sufficiently high. This is possible when favorable pairwise solute-solute interactions are substantial (here $\mathcal{E}/RT = -3.6$) so that bigger clusters are still energetically more favorable than smaller ones. Clustering of hydrophobic solutes in concentrated aqueous solutions has been attributed to their “cooperative” multiple-body potentials of mean force,^{12,18–20} and Tsai *et al.*¹² have discussed the potential relevance of this idea to the minimum hydrophobic core sizes in proteins. But Figs. 1–3 show that cooperative interactions are not necessary for solute aggregation. In fact, recent simulations by Rank and Baker²¹ indicate that contact interactions among multiple hydrophobic solutes may be slightly less favorable than the corresponding pairwise sums. In other words, there may be a small anti-cooperative effect in hydrophobic clustering, though this finding has recently been disputed.²⁹ (See also Ref. 30.)

Notwithstanding their apparent contradiction, the results of Tsai *et al.*¹² and those of Rank and Baker²¹ are not necessarily inconsistent. This is because the definition of “cooperativity” by Tsai *et al.* does not refer to deviations from either of the two meanings of additivity described above. By “cooperativity,” Tsai *et al.* mean that the interaction between two larger multiple-solute hydrophobic clusters is more favorable than that between two smaller clusters.¹² We do not recommend their usage of the term “cooperativity” in analyses of solute aggregation and protein folding. The reason is that the “cooperativity” of Tsai *et al.* can be consistent with the above-defined additivity (noncooperativity) or even anti-cooperativity of the underlying microscopic interactions. Physically, the association of bigger hydrophobic clusters would most likely involve larger buried hydrophobic surfaces and therefore more favorable as a whole than the association of smaller clusters, even if the underlying interactions themselves are additive or slightly anti-cooperative in terms of either (i) solute-solute interactions or (ii) some kind of surface area measure (see above). Therefore, for solutes with favorable pairwise interactions, the “cooperativity” of Tsai *et al.* is expected to almost always hold except in cases with very strong anti-cooperative effects ($\mathcal{E}_{\text{coop}} \gg 0$ in the present model, for example). Since the “cooperativity” of Tsai *et al.* holds in almost all systems, its usage is less informative than the definition of cooperativity adopted here. In fact, the present definition of interaction cooperativity is one that coincides with the commonly accepted usage in the protein folding theory context.^{29,42–44}

IV. CONNECTIVITY CONSTRAINTS ON CHAIN COLLAPSE

Aggregation of hydrophobic solutes is often used to model and rationalize hydrophobic collapse in protein folding.¹² An obvious difference between the two kinds of systems is that while solutes can disperse freely in solutions,

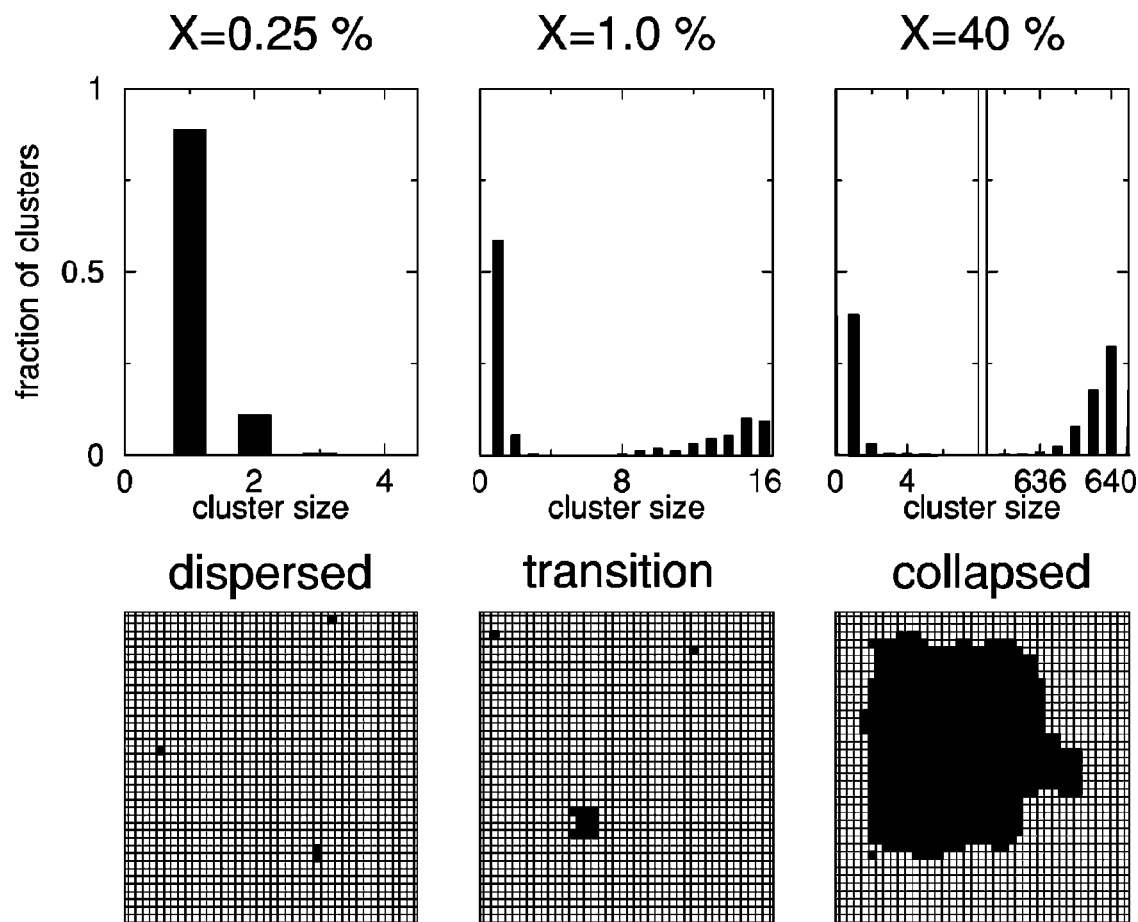


FIG. 2. Solute aggregation is possible with noncooperative intersolute interactions at high concentration. Results are from Monte Carlo sampling of the lattice solution model at $\mathcal{E}/RT = -3.6$ and $\mathcal{E}_{\text{coop}} = 0$ with (from left to right) 4, 16, and 640 solutes for three different solute mole fractions (X). Top: Solute cluster size distributions. (Left) The solutes are well dispersed in solution at small X . (Middle) At an intermediate X , the distribution is broad and bimodal, signaling a transition from the dispersed to the collapsed phase. (Right) At high X , a minority of the solutes are dispersed, while an overwhelming majority of them collapse to form one or a few big aggregates. (cf. Fig. 1). The $X = 40\%$ distribution at these two extremes is shown in two histograms. Cluster sizes outside their scales were not encountered in our simulations. Bottom: Snapshots of model solutions at the three X 's. Solutes and solvents are depicted, respectively, by filled (black) and open (white) small squares.

relative positions of the monomers of a protein are severely restricted because the monomers are covalently connected. Therefore, to better understand the relationship between the two processes, one needs to gain insight into the constraints imposed by chain connectivity on solvophobic association. Here we compare the collapse of an isolated chain with aggregation (as modeled earlier) by considering both processes in a common lattice framework.

We use the standard HP (hydrophobic/polar) lattice copolymer chain model.^{45,46} Each nearest-neighbor contact between two H monomers not consecutive along the sequence (i.e., an HH contact⁴⁵) is assigned an energy \mathcal{E} . All other contacts, including “connected contacts” between two sequentially consecutive H monomers, are neutral (have zero energy). The HP model interaction is additive. The energy of a conformation with h HH contacts is equal to $\mathcal{E}h$. We consider all $2^{18} = 262\,144$ HP sequences with $N_{\text{chain}} = 18$ monomers, 6349 of which have a unique lowest-energy conformation and are often used as model proteins.^{45–47} The exact density of states $g(h)$, which gives the number of conformations of a sequence with any given number of HH contacts, is determined for all 2^{18} sequences by enumerating the

5 808 335 possible conformations for each sequence. Only a small fraction of these densities have been determined previously.⁴⁶

For comparisons with solute aggregation, we may identify an H monomer with a solvophobic solute in the aggregation model above, and a P monomer as having the same properties as a model solvent molecule. Hence the HP sequence as a whole may be envisioned to be solvated also by the same solvent as in the solute aggregation model. To monitor the association of the n_{H} H monomers in a HP sequence, two collapse order parameters similar to Eq. (7) are defined, the first of which is

$$\alpha_{\text{chain}}^0 \equiv \frac{\langle h \rangle}{n_{\text{H}}}, \quad (8)$$

where the Boltzmann average of h is computed using the partition function $\sum_h g(h) \exp(-\mathcal{E}h/RT)$. α_{chain}^0 counts only HH contacts that favor collapsed conformations in the HP model. It does not count the number of instances that two H's reside next to each other along the sequence. We include these “connected contacts” in a different parameter,

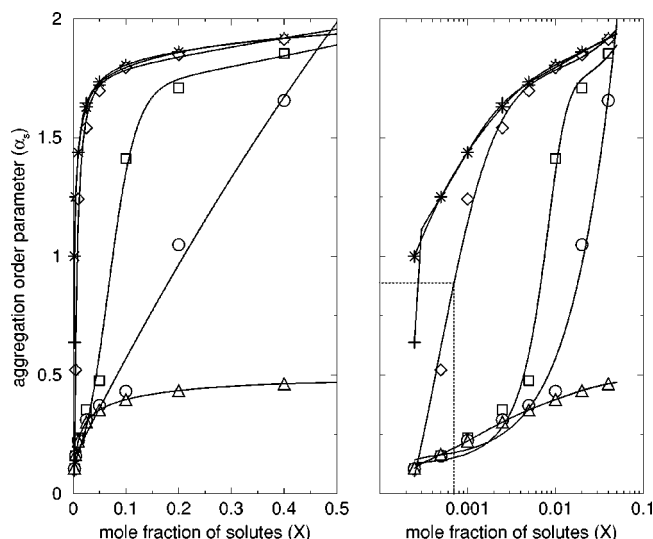


FIG. 3. Effects of cooperativity and anti-cooperativity on solvophobic aggregation. The format is the same as that in Fig. 1. Data points are from Monte Carlo simulations, all with $\mathcal{E}/RT = -3.6$, whereas $\mathcal{E}_{\text{coop}}$ equals (from top to bottom) -3.6 (asterisks), -2.4 (crosses), 0 (diamonds), $+2.4$ (squares), $+3.6$ (circles), and $+7.2$ (triangles). Solid lines and curves are mere guides for the eye. The critical mole fraction $X = \rho_a = -1/2B_2^*$ predicted using the second osmotic coefficient is shown for the $\mathcal{E}_{\text{coop}} = 0$ curve (dotted lines). Results in this plot clearly indicate that a negative $\mathcal{E}_{\text{coop}}$ can lead to a reduction of the critical mole fraction. However, this feature is not captured by the second osmotic virial coefficient because B_2^* does not account for three- and higher-body cooperative effects.

$$\alpha_{\text{chain}} \equiv \alpha_{\text{chain}}^0 + h_c/n_H, \quad (9)$$

where h_c is the number of sequential consecutive HH pairs along the given HP sequence. The motivation for considering α_{chain} is that it is directly comparable to α_s in Eq. (7) because they both measure the same geometrical property dictated by the relative positions of solvophobic chain monomers or independent solvophobic solutes.

We now compare how the solvophobic/hydrophobic compositions affect solute aggregation versus chain collapse.⁴⁸ Figure 4 shows a monotonic increase in average hydrophobic clustering (increasing α_{chain}) among the HP sequences as their hydrophobic compositions n_H/N_{chain} are increased. Hydrophobic clustering in HP sequences is concomitant with chain collapse.⁴⁵ The general trend in Fig. 4 is similar to that in Fig. 1 for independent solvophobic solutes, which show increasing aggregation with increasing concentration. The reason for this similarity is intuitive. Since connectivity localizes the chain monomers within a certain effective “volume,” a sequence with a higher hydrophobic composition has a higher “effective concentration”^{12,49} of hydrophobic monomers. Therefore their clustering properties resemble that of a solution with a higher solvophobic solute concentration.

However, there are also significant differences between solute aggregation and chain collapse. Figure 4 shows that as chain hydrophobic composition n_H/N_{chain} is increased, the degree of collapse as measured by α_{chain} does not exhibit any sharp change with respect to n_H/N_{chain} . The transition between a more “open” and a more “collapsed” composition regime is gradual. In contrast, for solvophobic aggregation,

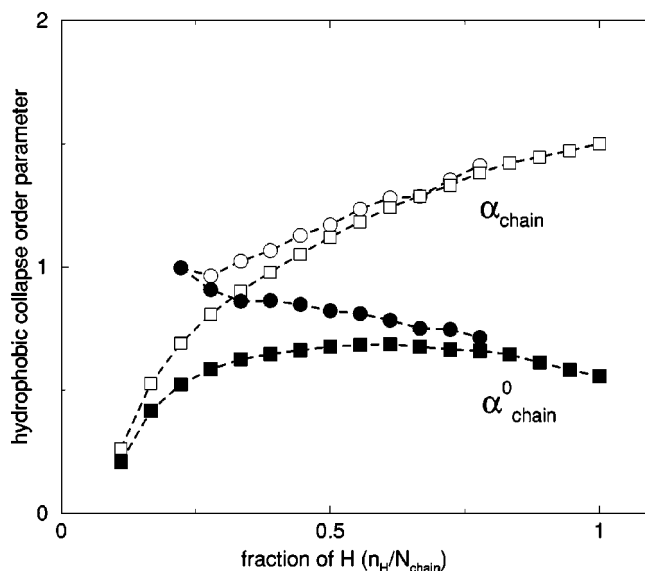


FIG. 4. Hydrophobic clustering among $N_{\text{chain}} = 18$ two-dimensional HP sequences at $\mathcal{E}/RT = -10.0$. All 2^{18} sequences are considered. The averages of α_{chain} and α_{chain}^0 over all sequences (squares) and unique sequences (circle) of given hydrophobic compositions n_H/N_{chain} are plotted by open and filled symbols, respectively. The dashed lines are guides for the eye.

sharp changes in α_s with respect to solute mole fraction X are observed even when the strength of the pairwise favorable solute-solute interaction (Fig. 1, $\mathcal{E}/RT < -2.5$) is less than a third of that between hydrophobic chain monomers in Fig. 4 ($\mathcal{E}/RT = -10.0$). For binary solutions, when the solvophobic attraction between solutes is sufficiently strong, there is a relatively well-defined critical X value demarcating mole fractions into a “dispersed” regime and an “aggregated” regime. Such sharp transitions imply that a small change in X in the critical region can lead to a huge change in the configurational entropy of the system. Obviously, the configurational freedom of a collection of solvophobic chemical groups is much more limited when they are linked monomers along a chain than when they are detached independent solutes. Because the maximum configurational entropy of HP sequences is intrinsically limited by chain connectivity, they show no sharp variation in sequence-averaged hydrophobic clustering as a function of hydrophobic composition.

Chain connectivity also imposes constraints on the packing of the hydrophobic monomers. Hence the maximum number of hydrophobic contacts a specific HP chain can form is sequence dependent and cannot be deduced from its hydrophobic composition alone. Figure 4 shows that for a given hydrophobic composition, sequences with unique ground-state conformations are on average more efficient in forming hydrophobic contacts. This effect is especially dramatic for α_{chain}^0 at low hydrophobic compositions. Only sequences with $n_H = 4-14$ can be unique in the present model. Hydrophobic packing efficiencies of example HP sequences are shown in Fig. 5.

Thus there are similarities as well as significant differences between solute aggregation and chain collapse, even when both processes are driven by the same underlying sol-

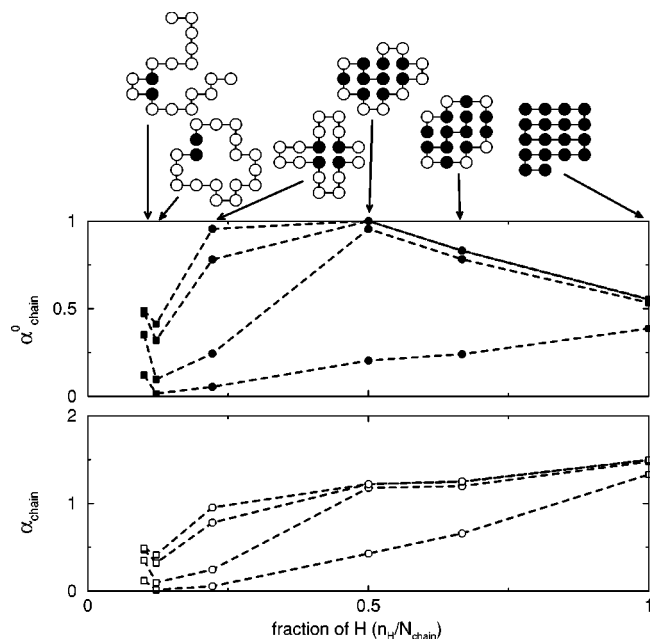


FIG. 5. Sequence effects in hydrophobic clustering: α_{chain}^0 (upper plot) and α_{chain} (lower plot) of six example HP sequences with (from top to bottom) $\mathcal{E}/RT = -6.0, -5.0, -3.0,$ and -1.0 are plotted. The dashed lines are only guides for the eye. At the top of the figure, each sequence is shown in either its unique ground-state conformation or one of its multiple ground-state conformations. H and P monomers are represented by filled and open circles respectively. Three of the sequences shown, with $n_H/N_{\text{chain}} = \frac{4}{18}, \frac{9}{18},$ and $\frac{12}{18}$, are unique. Others are not. Data for the two sequences with $n_H/N_{\text{chain}} = \frac{2}{18}$ are plotted with a small horizontal offset for clarity.

vophobic interaction. This underscores the indispensability of self-contained polymer models^{42,43} in the study of protein energetics, and cautions against drawing simplistic connections between sizes of solute aggregates and protein hydrophobic core volumes. Nevertheless, the same basic principle of competition between configurational entropy [arising from $g_s(h)$ or $g(h)$] and favorable solvophobic interactions applies to both processes. Aggregation or chain collapse occurs when the latter dominates over the former.

V. SUMMARY

Using a simple model, we have shown that although multiple-solute interaction cooperativity can enhance the thermodynamic stability of the aggregated state, they are not necessary for aggregation. Aggregation is possible with additive or even anti-cooperative interactions, provided that the anti-cooperativity is not too strong. The degree of aggregation increases with solute concentration. The calculations presented argue against a recent claim⁵⁰ that aggregation is constrained to occur at extreme low solute concentrations because of a presumed upper bound on excess free energy.⁵¹ We have also investigated the relationship between chain collapse and solute aggregation, an issue that is of relevance to protein folding.¹² We found that the two processes can shed light on each other, but there are significant quantitative differences between them because of the severe constraints imposed by chain connectivity on configurational freedom and on the packing of hydrophobic cores.

ACKNOWLEDGMENTS

We thank Danny Heap for his effort in maintaining our computing system. This work was supported in part by the Connaught Fund, a Premier's Research Excellence Award (Ontario) and Medical Research Council of Canada Grant No. MT-15323. One of the authors (H.S.C.) is a Canada Research Chair in Biochemistry.

- ¹ W. Kauzmann, *Adv. Protein Chem.* **14**, 1 (1959).
- ² C. Tanford, *The Hydrophobic Effect: Formation of Micelles and Biological Membranes* (Wiley, New York, 1980).
- ³ K. A. Dill, *Biochemistry* **29**, 7133 (1990).
- ⁴ B. Honig and A.-S. Yang, *Adv. Protein Chem.* **46**, 27 (1995).
- ⁵ See, e.g., (a) P. M. Harrison, H. S. Chan, S. B. Prusiner, and F. E. Cohen, *J. Mol. Biol.* **286**, 593 (1999); (b) *Protein Sci.* **10**, 819 (2001).
- ⁶ T. L. Hill, *Statistical Mechanics—Principles and Selected Applications* (McGraw-Hill, New York, 1956), p. 193.
- ⁷ T. L. Hill, *An Introduction to Statistical Thermodynamics* (Dover, New York, 1986), Chap. 20, pp. 371–397.
- ⁸ (a) R. Heyrovská, *Mar. Chem.* **70**, 49 (2000); (b) *J. Mol. Liq.* **81**, 83 (1999); (c) *J. Electrochem. Soc.* **143**, 1789 (1996).
- ⁹ (a) Y. V. Kalyuzhnyi, M. F. Holovko, and A. D. J. Haymet, *J. Chem. Phys.* **95**, 9151 (1991); (b) D. E. Smith, Y. V. Kalyuzhnyi, and A. D. J. Haymet, *ibid.* **95**, 9165 (1991); (c) J. Wang and A. D. J. Haymet, *ibid.* **100**, 3767 (1994).
- ¹⁰ D. C. Rapaport and H. A. Scheraga, *J. Phys. Chem.* **86**, 873 (1982).
- ¹¹ K. Watanabe and H. C. Andersen, *J. Phys. Chem.* **90**, 795 (1986).
- ¹² S. Tsai, M. Gerstein, and M. Levitt, *Protein Sci.* **6**, 2606 (1997).
- ¹³ R. H. Wood and P. T. Thompson, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 946 (1990).
- ¹⁴ W. Blokzijl and J. B. F. N. Engberts, *Angew. Chem. Int. Ed. Engl.* **32**, 1545 (1993).
- ¹⁵ P. T. Thompson, C. B. Davis, and R. H. Wood, *J. Phys. Chem.* **92**, 6386 (1988).
- ¹⁶ R. P. Kennan and G. L. Pollack, *J. Chem. Phys.* **93**, 2724 (1990).
- ¹⁷ A. Laaksonen and P. Stilbs, *Mol. Phys.* **74**, 747 (1991).
- ¹⁸ A. Wallqvist, *J. Phys. Chem.* **95**, 8921 (1991).
- ¹⁹ A. Wallqvist, *Chem. Phys. Lett.* **182**, 237 (1991).
- ²⁰ A. Wallqvist, *J. Chem. Phys.* **96**, 1655 (1992).
- ²¹ J. A. Rank and D. Baker, *Protein Sci.* **6**, 347 (1997).
- ²² S. Lüdemann, R. Abseher, H. Schreiber, and O. Steinhauser, *J. Am. Chem. Soc.* **119**, 4206 (1997).
- ²³ J. J. Kozak, W. S. Knight, and W. Kauzmann, *J. Chem. Phys.* **48**, 675 (1968).
- ²⁴ (a) H. S. Chan and K. A. Dill, *Annu. Rev. Biophys. Biomol. Struct.* **26**, 425 (1997); (b) D. R. DeVido, J. G. Dorsey, H. S. Chan, and K. A. Dill, *J. Phys. Chem. B* **102**, 7272 (1998), and references therein.
- ²⁵ S. Shimizu, M. Ikeguchi, S. Nakamura, and K. Shimizu, *J. Chem. Phys.* **110**, 2971 (1999).
- ²⁶ R. Brem, H. S. Chan, and K. A. Dill, *J. Phys. Chem. B* **104**, 7471 (2000).
- ²⁷ (a) S. Shimizu and H. S. Chan, *J. Chem. Phys.* **113**, 4683 (2000); (b) *J. Am. Chem. Soc.* **123**, 2083 (2001).
- ²⁸ V. Martorana, D. Bulone, P. L. San Biagio, M. B. Palma-Vittorelli, and M. U. Palma, *Biophys. J.* **73**, 31 (1997).
- ²⁹ C. Zaplewski, S. Rodziewicz-Motowidlo, A. Liwo, D. R. Ripoll, R. J. Wawak, and H. A. Scheraga, *Protein Sci.* **9**, 1235 (2000).
- ³⁰ S. Shimizu and H. S. Chan, *J. Chem. Phys.* **115**, 1414 (2001).
- ³¹ (a) M. S. Wertheim, *J. Stat. Phys.* **35**, 19 (1984); (b) **35**, 35 (1984); (c) **42**, 459 (1984); (d) **42**, 477 (1984); (e) P. H. Fries and J. Richardi, *J. Chem. Phys.* **113**, 9169 (2000).
- ³² (a) A. A. H. Padua and J. P. M. Trusler, *J. Chem. Phys.* **105**, 5956 (1996); (b) J. A. Anta, E. Lomba, and M. Lombardero, *Phys. Rev. E* **55**, 2707 (1997); (c) J. M. Bomont, N. Jakse, and J. L. Bretonnet, *Phys. Rev. B* **57**, 10217 (1998); (d) A. A. H. Padua and J. P. M. Trusler, *J. Chem. Phys.* **113**, 312 (2000); (e) N. Jakse, J. M. Bomont, I. Charpentier, and J. L. Bretonnet, *Phys. Rev. E* **62**, 3671 (2000).
- ³³ (a) L. A. Rodriguez-Guadarama, S. K. Talsania, K. K. Mohanty, and R. Rajagopalan, *J. Colloid Interface Sci.* **224**, 188 (2000); (b) E. La Nave, S. Sastry, F. Sciortino, and P. Tartaglia, *Phys. Rev. E* **59**, 6348 (1999); (c) W. Kob, *J. Phys.: Condens. Matter* **11**, R85 (1998).
- ³⁴ P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, NY, 1953).

- ³⁵W. G. McMillan and J. E. Mayer, *J. Chem. Phys.* **13**, 276 (1945).
- ³⁶T. L. Hill, *An Introduction to Statistical Thermodynamics* (Dover, New York, 1986), p. 341.
- ³⁷R. K. Pathria, *Statistical Mechanics* (Pergamon, New York, 1980), p. 262.
- ³⁸I. Prigogine and R. Defay, *Chemical Thermodynamics* (Longman, London, 1967), p. 221.
- ³⁹F. M. Richards, *Annu. Rev. Biophys. Bioeng.* **6**, 151 (1977).
- ⁴⁰K. Binder and M. Müller, *Int. J. Mod. Phys. C* **11**, 1093 (2000).
- ⁴¹N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- ⁴²H. S. Chan, *Proteins: Struct., Funct., Genet.* **40**, 543 (2000).
- ⁴³(a) H. Kaya and H. S. Chan, *Proteins: Struct., Funct., Genet.* **40**, 637 (2000); (b) *Phys. Rev. Lett.* **85**, 4823 (2000).
- ⁴⁴(a) A. Kolinski, W. Galazka, and J. Skolnick, *Proteins: Struct., Funct., Genet.* **26**, 271 (1996); (b) M.-H. Hao and H. A. Scheraga, *J. Chem. Phys.* **107**, 8089 (1997); (c) S. Takada, Z. Luthey-Schulten, and P. G. Wolynes, *ibid.* **110**, 11616 (1999); (d) M. P. Eastwood and P. G. Wolynes, *ibid.* **114**, 4702 (2001).
- ⁴⁵H. S. Chan and K. A. Dill, *Proteins: Struct., Funct., Genet.* **24**, 335 (1996).
- ⁴⁶E. Bornberg-Bauer and H. S. Chan, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10689 (1999).
- ⁴⁷A. Irbäck and E. Sandelin, *Biophys. J.* **79**, 2252 (2000).
- ⁴⁸The principal focus of the present lattice study is on solvophobic/hydrophobic composition. Here we do not compare transition sharpness of chain collapse versus solute aggregation as a function of temperature or interaction strength. Thermodynamic properties of HP model sequences as functions of temperature and sticking energy are summarized, e.g., in H. S. Chan, S. Bromberg, and K. A. Dill, *Philos. Trans. R. Soc. London, Ser. B* **348**, 61 (1995); 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); and Ref. 42.
- ⁴⁹T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2nd ed. (Freeman, New York, 1993), p. 163.
- ⁵⁰A. Marmur, *J. Am. Chem. Soc.* **122**, 2120 (2000).
- ⁵¹The premise of Marmur's claim is that the actual excess free energy of solvation cannot exceed RT , in which case if the mole fraction of solute i in a saturated solution $X_i < 1/e = 0.3679 \dots$ and therefore $-RT \ln X_i > RT$, it is impossible that the solutes are dispersed as individual molecules and formation of aggregate is the only way to satisfy the excess free energy upper bound. However, the general "order-of-magnitude" mathematical argument in Ref. 50 that purports to establish this upper bound is not valid. Consider the illustrative example whereby the partial molar volumes of the solute in a solution phase (\bar{V}_i) and in its pure liquid phase (V_i) are given by the van der Waals equation of state $(P + a/V^2) \times (V - b) = RT$, with $a = 9.22$, $b = 4.43$ for $\bar{V}_i(P)$ and $a = 25.4$, $b = 7.90$ for $V_i(P)$. Then for a liquidlike pressure $P = 0.0292$ ($RT = 0.594$), the excess free energy calculated by equation 1 in Ref. 50 is equal to $\int_0^P dP' [\bar{V}_i(P') - V_i(P')] = 1.70RT$, exceeding the claimed upper bound of RT . We note that similar $V(P)$ behaviors are readily found for real solution systems such as hexadecane in water. In the numerical example RT is in units of kcal/mol so $T = 299$ K. V and b are in liter/mol, a is in liter-kcal/mol², and P is in kcal/liter so $P = 1.2$ atm in the example. The example remains valid if the liter units are scaled with any dimensionless numerical factor.