

## Simple model of sickle hemoglobin

Andrey Shirayev,<sup>a)</sup> Xiaofei Li, and J. D. Gunton<sup>b)</sup>

*Department of Physics, Lehigh University, Bethlehem, Pennsylvania 18015*

(Received 24 October 2005; accepted 15 May 2006; published online 11 July 2006)

A microscopic model is proposed for the interactions between sickle hemoglobin molecules based on information from the protein data bank. A solution of this model, however, requires accurate estimates of the interaction parameters which are currently unavailable. Therefore, as a first step toward a molecular understanding of the nucleation mechanisms in sickle hemoglobin, a Monte Carlo simulation of a simplified two patch model is carried out. A gradual transition from monomers to one dimensional chains is observed as one varies the density of molecules at fixed temperature, somewhat similar to the transition from monomers to polymer fibers in sickle hemoglobin molecules in solution. An observed competition between chain formation and crystallization for the model is also discussed. The results of the simulation of the equation of state are shown to be in excellent agreement with a theory for a model of globular proteins, for the case of two interacting sites.

© 2006 American Institute of Physics. [DOI: [10.1063/1.2210935](https://doi.org/10.1063/1.2210935)]

### I. INTRODUCTION

The condensation of globular proteins from solution is a subject of considerable experimental and theoretical activity. One one hand, it is important to grow high quality protein crystals from solution in order to be able to determine protein structure (and thus function) from x-ray crystallography. On the other hand, many diseases are known to be caused by undesired condensation of proteins from solution. In both cases one needs to have a reasonably detailed model of the protein-protein interactions in solution in order to predict the phase diagram, condensation rate, and growth kinetics. This is a major challenge to theorists, as these protein-protein interactions arise from many sources and are still relatively unknown in most cases.

An important example of undesired protein condensation occurs with sickle hemoglobin (HbS) molecules in solution. It is known that deoxygenated sickle hemoglobin molecules in red blood cells can undergo a polymerization process that involves a two step nucleation mechanism; this leads to the formation of polymer fibers in the cell.<sup>1,2</sup> These fibers distort the cells and make it difficult for them to pass through the capillary system on their return to the lung. A direct determination of the homogeneous nucleation of HbS fibers *in vitro* has shown that the nucleation rates are of the order of  $10^6$ – $10^8$  cm<sup>-3</sup> s<sup>-1</sup> and that the induction times agree with Zeldovich's theory.<sup>3</sup> These rates are comparable to those leading to erythrocyte sickling *in vivo*. They are also approximately nine to ten orders of magnitude larger than those known for other protein crystal nucleation rates, such as lysozyme.

Consequently a goal of current research is to understand at a molecular level this nucleation process and by controlling the conditions on which the nucleation depends, to slow down the nucleation rate such as to prevent the polymeriza-

tion from occurring while HbS is in its deoxygenated state in the cells. To do this requires understanding the protein-protein interactions, in order to predict the phase diagram and nucleation rate for sickle hemoglobin molecules. The phase diagram for HbS is only partially known experimentally. It is known that there is a solubility line separating monomers and fibers;<sup>4,5</sup> there is also evidence for a spinodal curve with a lower critical point.<sup>6</sup> In a previous publication<sup>7</sup> Shirayev *et al.* obtained a phase diagram that was qualitatively similar to the experimental situation, namely, liquid-liquid phase separation and spinodal curves with a lower critical point. In addition, we determined the location of the liquids and crystallization lines for the model, as shown in Fig. 1.

However, although yielding a lower solution critical point, this model was unable to predict the formation of polymer fibers, as it was based on a spatially isotropic, short range protein-protein interaction (e.g., a square well or a modified Lennard-Jones potential energy). Fiber formation clearly requires anisotropic interactions. In this paper we propose an anisotropic model for the HbS–HbS interactions, based on an analysis of the contacts for HbS crystals from the protein data bank. We also define an order parameter to describe the polymerization of this model. The full model is rather complex. More importantly, it involves several interaction parameters whose values are currently unknown. Therefore, as a first step in a longer term effort to obtain an accurate, minimalist microscopic model of HbS polymerization, we study a simplified version of our model (a two patch model). Our primary goal here is to gain some insight into the nature of the chain formation for this model by studying the free energy barrier associated with the phase transition, similar to studies performed for the fluid-fluid and fluid-solid transitions in other systems by Frenkel and co-workers.<sup>8–10</sup> We do not study the kinetics of the actual polymerization process, which is known to be important for HbS polymerization, as the model is not sufficiently realistic to warrant

<sup>a)</sup>Electronic mail: [andrey@lehigh.edu](mailto:andrey@lehigh.edu)

<sup>b)</sup>Electronic mail: [jdg4@lehigh.edu](mailto:jdg4@lehigh.edu)

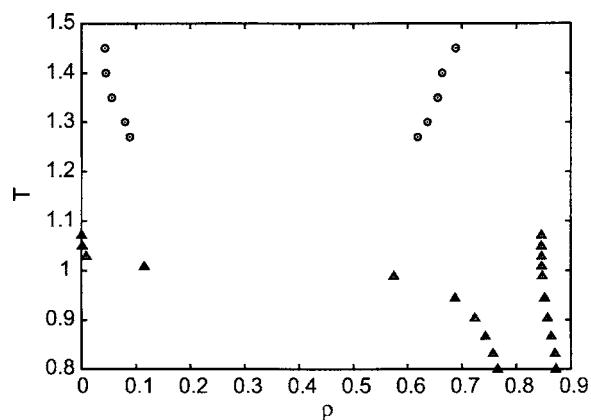


FIG. 1. Phase diagram of a modified Lennard-Jones model including solvent-solute interactions. Open triangles denote liquid-solidus line; open circles denote fluid-fluid coexistence. From Ref. 7. Details of the model are given in this reference.

this. This of course means that we cannot study nonequilibrium aspects of the problem, other than possible nucleation mechanisms that would be observed from determining nucleation free energy barriers. We determine some aspects of the phase diagram for the two patch model via standard Monte Carlo simulation and biasing techniques<sup>8-10</sup> and show, in particular, that the model yields one dimensional chains that are somewhat similar to the polymer fibers formed in HbS nucleation. These chains arise from the simple two patch anisotropy of the model. Real HbS fibers, however, arise from more complex interactions and have a diameter of about 21 nm. In addition, the strands within the fiber are packed into double strands. Thus the two patch model is too simple to describe the polymer fiber phase transition observed in HbS. Future work will be necessary to obtain reasonable estimates of the interaction parameters in the full model, in order to obtain an accurate microscopic model for the polymer fiber formation.

The outline of the paper is as follows. In Sec. II we propose an anisotropic interaction model for the pair interactions between HbS molecules. In Sec. III we define an order parameter that measures the degree of polymerization in the system. In Sec. IV we present the results of our Monte Carlo simulation for a two patch approximation to the full model, since we are unable to make realistic estimates for the interaction parameters for the full model. A biasing technique is used in order to examine the free energy barrier associated with the transition. We also summarize the results of a perturbation theory as applied to an eight patch model and to our two patch model. In the latter case we show that the simulation results are in excellent agreement with this theory. In Sec. V we present a brief conclusion.

## II. ANISOTROPIC MODEL FOR THE HEMOGLOBIN S POLYMERIZATION

Protein molecules in general, and sickle hemoglobin molecules, in particular, are very complicated objects, typically consisting of thousands of atoms. There are many types of forces between protein molecules in solution, such as Coulomb forces, van der Waals forces, hydrophobic interac-

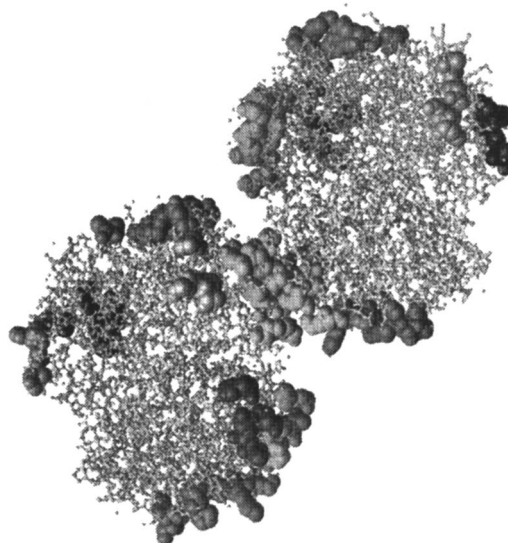


FIG. 2. Location of the residues participating in different contacts. Yellow areas denote the lateral contacts, and green areas denote the axial contacts.

tions, ion dispersion forces, and hydrogen bonding. Although these interactions are complex, considerable success in predicting the phase diagrams of several globular proteins in solution has been accomplished by using rather simple models with spatially isotropic interactions. These models share in common a hard core repulsion together with a short range attractive interaction (i.e., the range of attraction is small compared to the protein diameter). However, in general the protein-protein interactions are anisotropic, often arising from interactions between specific amino acid residues on the surfaces of the interacting molecules; i.e., certain areas of a protein surface interact with certain areas of another molecule's surface. Thus there has been some recent work attempting to model these anisotropic interactions. In such models a given protein molecule is represented by a hard sphere with a set of patches on its surface.<sup>11-16</sup> Intermolecular attraction is localized on these patches. Typically the models assume that two protein molecules interact only when they are within the range of attractive interaction and when the vector joining their centers intersects patches on the surface of both molecules. The fluid-fluid diagram for such a model was studied by Kern and Frenkel<sup>16</sup> in a Monte Carlo simulation and by Sear theoretically.<sup>14</sup> In these studies all patches are assumed to interact with each other equally. This approximation gives a good qualitative picture of the possible importance of anisotropy in protein phase diagrams. However, when we consider the behavior of a specific globular protein, this approximation is too simple and does not reflect the actual structure of protein aggregates. This is particularly important in the fiber formation of HbS molecules in solution. Figure 2 shows interacting HbS molecules, with the different pairs of interacting regions in the crystal state of HbS. We use this information, in accordance with the contact information in the HbS crystal,<sup>17,18</sup> to develop a model for the anisotropic interactions between the HbS molecules.

To develop an anisotropic model to describe these interacting molecules, we allow for the possibility of different interaction strengths between different pairs of these interact-

ing patches. To characterize such interactions, we introduce an interaction matrix  $\{\epsilon_{kl}\}_{mm}$ , where  $\epsilon_{ij}$  is the strength of interaction between the  $k$ th and  $l$ th patches and  $m$  is the total number of patches on a protein surface. This is a symmetric matrix that describes the strength of interaction between each pair of patches. In our particular model we choose a square well attraction between patches, although this is not a necessary restriction on the interactions. Thus the interaction matrix consists of the well depth values for the different patch-patch interactions. We define the pair potential between two molecules in this case in a way similar to Refs. 14 and 16, but generalizing to the case of different patch-patch interactions,

$$U_{i,j}(\mathbf{r}_{ij}, \Omega_i, \Omega_j) = U_{ij}^0(r_{ij}) \sum_{k,l} \epsilon_{kl} \Theta_k(\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{n}}_{ik}) \Theta_l(-\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{n}}_{jl}). \quad (1)$$

Here  $m$  is the number of patches,  $\Omega_i$  is the orientation (three Euler angles, for example) of the  $i$ th molecule,  $U_{ij}^0$  is the square well potential for the unit well depth, and  $\hat{\mathbf{n}}_{ik}$  is the  $k$ th patch direction in the laboratory reference of frame,

$$\hat{\mathbf{n}}_{ik} = \tilde{R}(\Omega_i) \hat{\mathbf{n}}_k^0. \quad (2)$$

Here  $\hat{\mathbf{n}}_k^0$  is the  $k$ th patch direction in the  $i$ th molecule reference of frame and  $\tilde{R}(\Omega_i)$  is a rotation matrix between the  $i$ th molecule and the laboratory reference of frame.  $\Theta(x)$  in (1) is a step function,

$$\Theta_k(x) = \begin{cases} 1, & \text{if } x \geq \cos \delta_k \\ 0, & \text{if } x < \cos \delta_k, \end{cases} \quad (3)$$

where  $\delta_i$  is the half open angle of the  $i$ th patch. In other words,  $\Theta(\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{n}}_{ik})$  is equal to 1 when the vector joining two molecules intersects the  $k$ th patch on the surface. If patches do not overlap then the sum in Eq. (1) has at most one term. The radial square well dependence with range  $\lambda$  is given by

$$U_{ij}^0(r) = \begin{cases} \infty, & \text{for } r < \sigma \\ -1, & \text{for } \sigma \leq r < \lambda\sigma \\ 0, & \text{for } r \geq \lambda\sigma. \end{cases}$$

Until now what we have done is valid for any pair of molecules interacting through pairwise patch interactions. We now consider the specific case of the HbS molecule. As noted above, in agreement with contact information for the HbS crystal,<sup>17,18</sup> this molecule has two lateral and two axial patches that are involved in intradouble strand contacts and four more patches involved in interdouble strand contacts. Thus we have eight possible patches for the HbS molecule (Fig. 2). One of the lateral patches contains a  $\beta 6$  valine residue and another lateral patch contains an acceptor pocket for this residue. But it is known<sup>5</sup> that only half the mutated sites are involved in the contacts in the HbS crystal. Thus we have another possible set of eight patches similar to the first one. The total number of patches is therefore sixteen (two equal sets of eight patches). The interaction matrix can be built assuming that the first lateral patch (of any set) can interact only with the second lateral patch (of any set). The same is true for axial patches. The remaining four patches in each set

can be divided into pairs in a similar way, in accordance with Ref. 17. This gives the following interaction matrix (for one set of eight patches):

$$Y = \begin{pmatrix} 0 & \epsilon_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \epsilon_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \epsilon_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \epsilon_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \epsilon_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & \epsilon_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \epsilon_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & \epsilon_4 & 0 \end{pmatrix}, \quad (4)$$

where  $\epsilon_1$  is the strength of the lateral contact,  $\epsilon_2$  is the strength of the axial contact, and  $\epsilon_3$  and  $\epsilon_4$  are the strengths of the interdouble strand contacts. We will refer to this model as the “full model” for HbS.

### III. POLYMERIZATION ORDER PARAMETER IN A SYSTEM OF PATCHY HARD SPHERES

One of the goals of a study of HbS molecules in solution is to calculate the free energy barrier that separates the monomer solution from the aggregate (polymer chains/fibers) state. For this purpose we have to specify an “order” parameter that measures the degree of polymerization in the system. The structure of the aggregate depends strongly on the configuration of patches. Therefore, to separate the aggregate state from the monomer state, the order parameter should reflect the configuration of patches. (Note that the order parameter as defined below is only zero in the case in which there are only monomers.) Since in our model the regions on the molecular surface not covered by patches do not interact (except through the hard core repulsion), we can measure the degree of polymerization by measuring the fraction of the patches involved in actual contacts.

We assume that any two particles at any given time can have no more than one contact between each other. This condition is a little stronger than just a nonoverlap of the patches. For each pair of particles we introduce a quantity that shows how much these particles are involved in polymerization (basically, showing the presence of the contact between them),

$$\psi_{ij}(\mathbf{r}_i, \mathbf{r}_j, \Omega_i, \Omega_j) = \sum_{k,l}^{N_p} w_{kl} f_k(\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{n}}_{ik}) f_l(-\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{n}}_{jl}), \quad (5)$$

where  $w_{kl}$  is a weight of the contact between the  $k$ th patch of the  $i$ th molecule and the  $l$ th patch of the  $j$ th molecule. We choose the weight matrix to be the interaction matrix.  $f_k(x)$  is equal to  $x$  for  $x > \cos \delta_k$  and is zero otherwise. Due to our assumption of only one contact per pair of particles, the sum in (5) has at most one nonzero term. We next define the order parameter of one particle to be

$$\psi_i(\mathbf{r}_i) = \frac{\sum_j \psi_{ij}}{\sum_{k,l} w_{kl}}. \quad (6)$$

The term in the denominator is a normalization constant. The order parameter of the whole system is

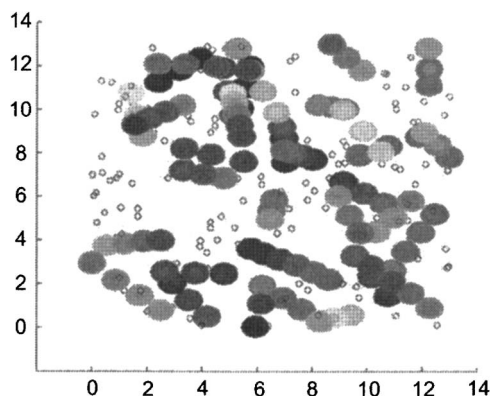


FIG. 3. Intermediate state of one dimensional chain formation using the two patch model. The actual particle size is equal to the size of the colored spheres. Particles that are not involved in any chains are scaled down and shown as small blue circles. Particles that are part of some chains are shown as colored spheres. Color shows the depth of the particle ( $z$  axis); blue is the deepest and red is the least deep. Simulations were performed at low temperature and high supersaturation.

$$\psi = \frac{1}{N} \sum_i^N \psi_i. \quad (7)$$

This choice of order parameter reflects the patch configuration; the magnitude of the order parameter increases as the number of contacts in the system increases. It is also rotationally invariant. However, this construction has its disadvantages. For some choices of weight matrices it is possible that a fewer number of contacts could lead to a larger order parameter if these contacts have significantly larger weights. However, the choice of the weight matrix equal to the interaction matrix seems to be natural.

#### IV. TWO PATCH MODEL AND ONE-DIMENSIONAL CHAIN FORMATION

The full model described earlier for the sickle hemoglobin molecule is complex and has several interaction parameters which have not yet been determined from experiment. Because we are interested in studying the fiber formation, we use a simplified model that still can produce fiber chains. We simplify the original model by reducing the number of contacts. Since one important feature of HbS fibers is the presence of twisted, quasi-one-dimensional chains, we consider a system of particles with only two (axial) patches. This model is obviously not an accurate description of interacting HbS molecules, but it can lead to the formation of one dimensional chains. The interaction matrix for the simplified two patch model is just

$$Y = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (8)$$

Since there are two patches on each sphere, there is only a fluid-solid phase transition. (The full model described above can have a fluid-fluid phase transition as well.) The formation of the one dimensional chains, therefore, is a gradual transition from the fluid phase as the density of the molecules is increased. Figure 3 shows chains that result in our simulation.

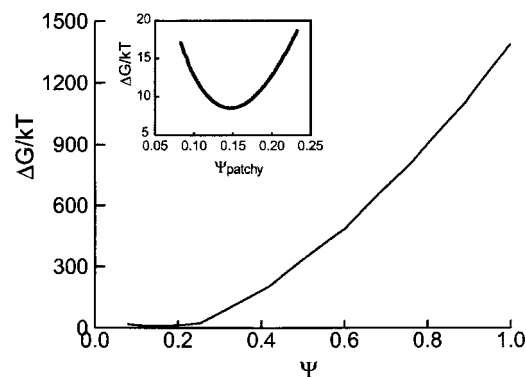


FIG. 4. Plot of the free energy  $\Delta G$  in units of  $kT$  vs the order parameter for the patches,  $\Psi$ , defined in the text. Simulations are at a temperature  $kT = 0.185$  and pressure  $P = 1.0$ . The minimum at  $\Psi$  around 0.15 corresponds to a liquid state which is a mixture of monomers and dimers.

Since the formation of these chains is a gradual transition as one increases the density, they do not arise from homogeneous nucleation. A necessary property of nucleation is the existence of a nucleation barrier in the free energy dependence on the order parameter. This barrier should separate two wells; one well corresponds to the metastable phase, the other to the more stable phase. In our case there will not be such a barrier. To examine the nature of the chain formation, we determined the dependence of the free energy on the patch order parameter by performing two series of umbrella sampling Monte Carlo simulations. The first set of simulations was done at  $T = 0.185$  and  $P = 1.0$ .

In this case the initial liquid state does not crystallize in the absence of the biasing and remains in the liquid state. The order parameter has only one minimum about  $\Psi_{\text{patch}} = \Psi \cong 0.15$ , corresponding to a mix of monomers and dimers (Fig. 4). As we increase the pressure to  $P = 1.6$  the free energy now has a minimum at a lower value of  $\Psi$  that corresponds to the liquid state (Fig. 5) and a second minimum at  $\Psi \approx 1$ . This second minimum corresponds to a crystal state with all the patches involved in contacts. Thus we see that this double-well free energy describes a liquid-solid phase

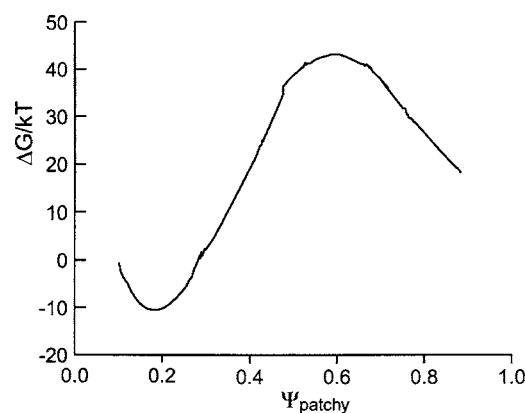


FIG. 5. Plot of the free energy  $\Delta G$  in units of  $kT$  vs the order parameter for the patches,  $\Psi$ , defined in the text. Simulations are at a temperature  $kT = 0.185$  and pressure  $P = 1.6$ . The minimum at  $\Psi$  around 0.2 corresponds to a mix of monomers and dimers, while the minimum at  $\Psi$  around 1 corresponds to a crystal state in which all the patches are in contact.

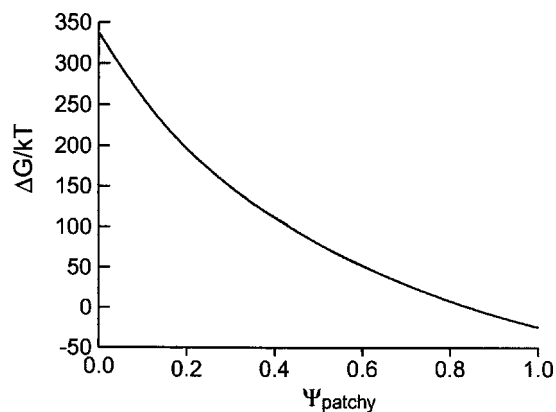


FIG. 6. Plot of the free energy  $\Delta G$  in units of  $kT$  vs the order parameter for the patches,  $\Psi$ , defined in the text. Simulations are at a temperature  $kT = 0.185$  and pressure  $P = 2.35$ . The minimum at  $\Psi$  in the vicinity of 1 is the crystal state.

transition rather than to a monomer- (one dimensional) fiber transition. This liquid-solid transition is what one would expect for this model.

Thus, as expected, there is no nucleation mechanism for the monomer-fiber transformation. In the case of  $T = 0.185$  and  $P = 2.35$  the system simultaneously crystallizes and increases its number of patch contacts. The system successfully reaches an equilibrium crystal state and the free energy has only one minimum at this state, as seen in Fig. 6. However, at lower temperature the picture is quite different. At  $T = 0.1$  and  $P = 0.01$  the fibers form before crystallization can occur. The free energy has one minimum at  $\Psi$  around 0.9, but the system is not crystallized. As the set of fibers is formed, the dynamics slows down significantly and the system becomes stuck in a nonequilibrium state. Figure 3 shows an example of a typical configuration for this nonequilibrium state, corresponding to a set of rods in a “glassy” state.

### A. Simulation details

The umbrella sampling simulations were performed on a system of  $N$  particles, with  $N = 500$ , in the  $N$ - $P$ - $T$  ensemble, with a range of interaction given by  $\lambda = 1.25$ . The equation of state simulations were also performed in the  $N$ - $P$ - $T$  ensemble with 500 particles. The details of the umbrella sampling technique can be found in Refs. 8 and 19. In short, this method is based on a biasing of the original interactive potential in such a way that the system is forced to reach otherwise inaccessible regions of the configuration space. In particular, for simulations starting in a liquid state, the system is forced towards larger values of the order parameter. Since the actual dependence of the free energy on the order parameter (for the entire range of values of  $\Psi$ ) is not known, the biasing function is chosen to be quadratic, i.e.,

$$U_{\text{biased}}(\mathbf{r}^N) = U_{\text{unbiased}}(\mathbf{r}^N) + k(\Psi(\mathbf{r}^N) - \Psi_0)^2,$$

where the parameters  $k$  and  $\Psi_0$  determine which region of values of  $\Psi$  would be sampled in the simulation. By changing these parameters we can sample the entire region of values of  $\Psi$ . The parameter  $k$  determines the width of the ranges of the order parameter that can be simulated: The higher the value of  $k$ , the narrower the region of order parameter values

that can be simulated in one run. The widest possible range is determined by the curvature of the free energy well obtained in unbiased simulations. At higher values of  $k$ , when the distribution is narrower, the system equilibrates faster. However, it takes more simulations to cover the whole region of order parameter values. Thus there is a trade off between the length of a single simulation and the number of simulations required to obtain the free energy at all values of the order parameter.<sup>8</sup> The parameter  $\Psi_0$  determines the shift of the maximum of the biased probability distribution from the unbiased value. The maximum of a final distribution will be somewhere between the  $\Psi_{\text{min}}$ —order parameter value at the bottom of the energy well and  $\Psi_0$ . When the unbiased distribution has two maxima (a two phase system), the biased distribution also has two maxima. The order parameter value for each of two maximums of the biased distribution is closer to  $\Psi_0$  than in the unbiased distribution.

Since the particle interactions are highly anisotropic, it takes longer to equilibrate them than for the case of particles with centrosymmetric interactions. For two particles to “bind together,” they have to not only get to be together, but also to be simultaneously oriented in the right way. This becomes a problem particularly at higher values of the order parameter and higher values of density, since more interparticle contacts are required. The equation of state calculations at lower densities required about  $10 \times 10^6$  steps to equilibrate and  $40 \times 10^6$  steps to obtain the statistics. At higher densities it took up to  $100 \times 10^6$  steps to equilibrate and  $200 \times 10^6$  more steps to obtain the statistics. For the umbrella sampling simulations the parameter  $k$  was set to  $k = 2000$ . This value is slightly higher than the curvature of the unbiased fluid free energy well. The number of steps required to equilibrate the biased system increases with increasing the biasing. It usually took about  $200 \times 10^6$ – $300 \times 10^6$  steps to equilibrate the system and another  $500 \times 10^6$ – $700 \times 10^6$  steps to get the statistics. Since in umbrella sampling simulations the calculation of the order parameter is required at every step, these simulations are much slower than the equation of state calculations. For the system of 500 particles it took about 20 hours per  $100 \times 10^6$  steps. Four to five simulations were needed to cover the required region of the order parameter. The statistical error of the Monte Carlo simulations is proportional to  $n^{-1/2}$ , where  $n$  denotes how many times the particular outcome was sampled. Around the maximum of the probability distribution this value was about  $n = 40\,000$ , which gives a statistical error less than 1%. At the edges of distribution, we used results with  $n$  not less than 50, which gives a statistical error of about 15%. The sampling error is much higher on the edges of the distribution. However, these are the regions used to fit two umbrella sampling runs. Thus if there is no significant overlap between two runs, the error involved in fitting of the results of these runs may be significant and the resulting free energy profile may be not smooth.<sup>20</sup> In our simulations two neighboring simulated regions of the order parameter did not quite overlap. To obtain an overlap, a quadratic extrapolation of the data was used. This was sufficiently accurate for our purpose, since the main goal was not to calculate the exact free energy barrier, but

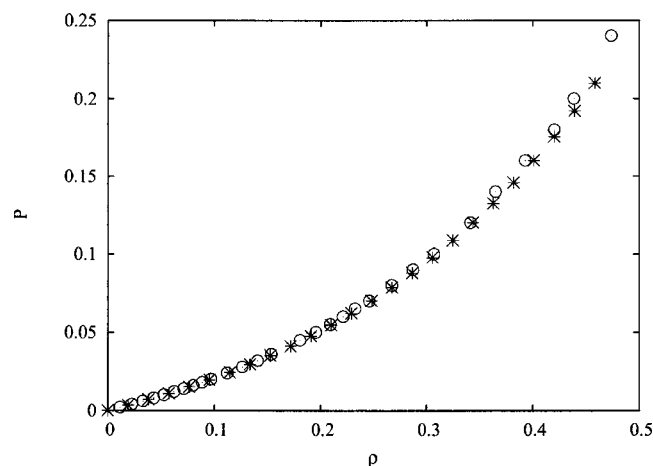


FIG. 7. A plot of the pressure  $p$  vs density,  $\rho$ , in the fluid phase at  $kT=0.17$  (in units of the well depth  $\epsilon$ ). The open circles are the results of the simulation. The asterisks are theoretical results obtained from Sear (Ref. 14). The number of patches  $m=2$ , with a patch angle of about  $\delta=52^\circ$  and a range of interaction  $\lambda=1.25$ , i.e.,  $r_c=1.25\sigma$ , where  $\sigma$  is the hard core diameter.

rather to determine whether the barrier exists. The existence of a barrier corresponds to two phases, one at each well minimum.

An interesting result of the simulations is that at intermediate pressures the system when biased to large values of  $\Psi$  starting from an initial liquid state has a very large volume. However, if the system is started from an initial (fcc) crystal state, the volume of the system remains small (still in a crystal state), while the order parameter value is around 1. This observation suggests that at not very high supersaturation the biased system started from the fluid ends up by forming a few long fibers, rather than a set of fibers that are packed into a crystal lattice. The fiber formation time therefore is much smaller than the crystallization time at low supersaturation. Starting from a fcc initial condition, however, the particles just reorient within the crystal lattice to form the fibers, remaining in the crystal state. At higher pressure, and therefore higher supersaturation, (such as in Fig. 5), the crystallization occurs in a time comparable with the fiber formation time. While the system is forced to form the fibers, these fibers pack into the crystal lattice.

## B. Equation of state

The equation of state for the two patch model is shown for the two low temperatures studied in Figs. 7 and 8.

Also shown in these figures are results from a theory for the  $m$  site model,<sup>12</sup> of globular proteins due to Sear.<sup>14</sup> This model is identical to our patch model defined in Sec. II for the case in which the various interaction parameters are equal. Sear studied the  $m$ -site model ( $m$  patches) using the Wertheim perturbation theory<sup>11</sup> for the fluid phase and a cell model for the solid phase<sup>21</sup> and showed that the model exhibits a fluid-fluid transition (for  $m > 2$ ) which is metastable with respect to the fluid-solid transition for most values of the model parameters. For  $m=2$ , however, there is only a fluid-solid transition. As can be seen from Fig. 7 and 8, the theory yields results which are in excellent agreement with

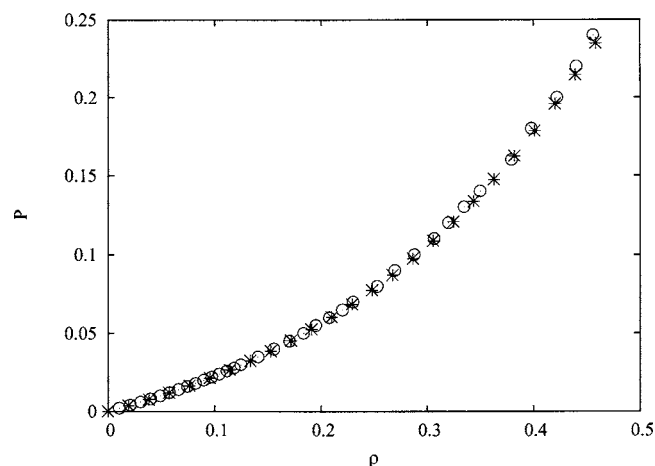


FIG. 8. A plot of the pressure vs density in the fluid phase at  $kT=0.185$ . The open circles are the results of the simulation. The asterisks are theoretical results obtained from Sear (Ref. 14). Same values for parameters as in Fig. 7.

the results of our simulation. For completeness we show the theoretical prediction for the fluid-solid transition for  $m=2$  in Fig. 9, assuming a fcc crystal structure.

We also show in Fig. 10 the results of the theory for the case  $m=8$ , as the model discussed in Sec. II has eight pairs of interacting patches. Since the interactions between the different sites in the model studied by Sear are assumed to be equal, the model lacks the anisotropy discussed in Sec. II that is necessary to account for the fiber formation in HbS molecules. Nevertheless, it is quite instructive to know the phase diagram for this case.

The fluid-fluid binodal curve has an upper critical point for  $m > 2$  (e.g., Fig. 10), unlike the case for HbS. In that case experimental measurements by San Biagio and Palma<sup>6</sup> display a spinodal curve. Such a spinodal implies the existence of a binodal curve with a lower critical point. However, as shown by Shirayayev *et al.*<sup>7</sup> the lower critical point reflects the crucial role of the solvent in the case of HbS in solution. The solvent is not taken into account in our model, but presumably if one would include a solute-solvent coupling simi-

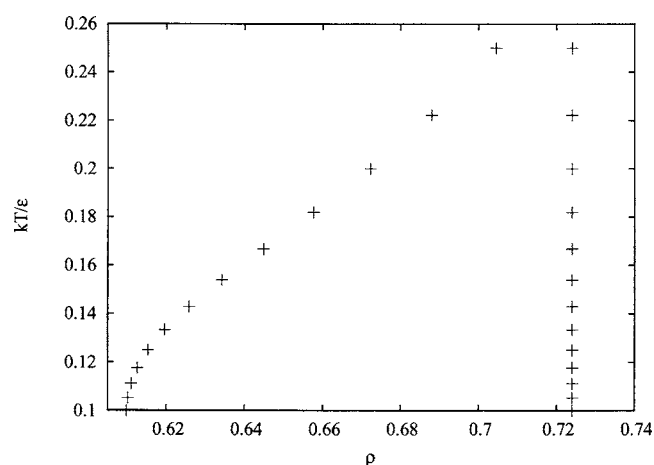


FIG. 9. Theoretical prediction for the phase diagram for the two patch model discussed in the text. The theory is due to Sear (Ref. 14). Same values for parameters as in Fig. 7.

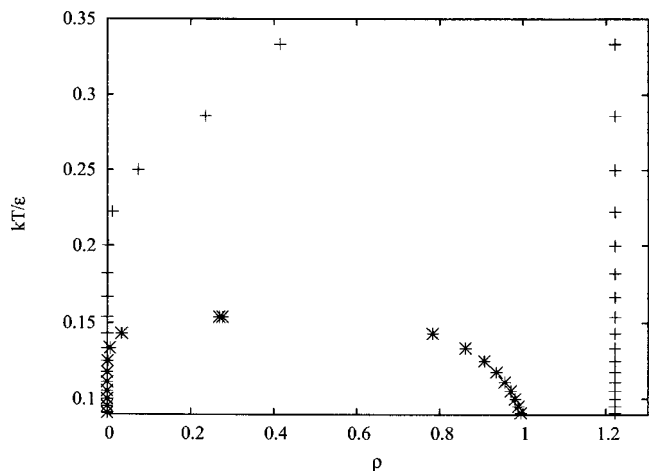


FIG. 10. Theoretical prediction for phase diagram for the model of HbS discussed in the text, in which all interaction parameters are equal. Theory is due to Sear (Ref. 14). Here  $m=8$ ,  $\lambda=1.05$ , and  $\delta=51^\circ$ . The fluid-fluid transition is metastable.

lar to that of Ref. 7, this coupling could change the phase diagram shown in Fig. 10 to one with a lower critical point, as found by Shirayev *et al.*,<sup>7</sup> e.g., Fig. 1.

Finally, we note that Jackson *et al.*<sup>12</sup> used Wertheim's theory for the two site model to predict that the fraction of molecules that are present in chains of length  $n$  is given by

$$nX^2(1-X)^{n-1}, \quad (9)$$

while the average chain length,  $\langle n \rangle$  is given by

$$\langle n \rangle = 1/X. \quad (10)$$

Here  $X$  is the fraction of sites that are not bonded to another site and is given by<sup>14</sup>

$$X = \frac{2}{1 + [1 + 4\rho K g_{hs}^c \exp(\beta\epsilon)]^{1/2}}, \quad (11)$$

where  $g_{hs}^c$  is the contact value of the pair distribution function for a fluid of hard spheres. The quantity  $K$  is given by the expression

$$K = \pi\sigma^2(r_c - \sigma)(1 - \cos(2\delta))^2. \quad (12)$$

For example, the theoretical prediction for the average chain length as a function of density (at  $kT=0.185$ ) is shown in Fig. 11. Also shown for comparison in Fig. 11 is an approximation for the average chain length obtained from our simulation results for the order parameter. Had we chosen to use a function  $f=1$  in our definition in Sec. III [rather than  $f(x)=x$  for  $x > \cos \delta$ ] the order parameter would have been equal to  $1-X$ , i.e., the fraction of sites that are bonded to another site. In that case the average chain length would be equal to  $1/(1-\Psi)$ . As shown in the figure,  $1/(1-\Psi)$  is in general less than the average chain length, due to our choice of  $f$  in Sec. II.

## V. CONCLUSION

In this paper we have carried out the first step in the process of developing an accurate microscopic model of sickle hemoglobin. We have used standard Monte Carlo

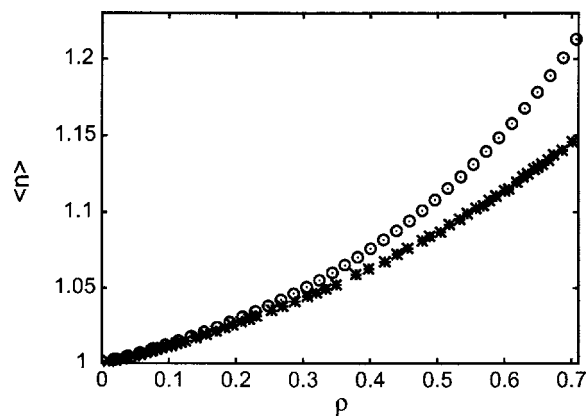


FIG. 11. Theoretical prediction for the average chain length as a function of density at  $kT=0.185$  (open circles). Also shown are the simulation results for  $1/(1-\Psi)$  (asterisks): see discussion in text.

simulation methods to study certain aspects of the properties of a simplified two patch interaction model. A gradual transition from monomers to one dimensional chains is observed as one varies the density of molecules at fixed temperature. This transition is a natural consequence of the two patch anisotropy of the model. An observed competition between chain formation and crystallization for the model is also discussed. The results of the simulation of the equation of state are shown to be in excellent agreement with a theory for a model of globular proteins, for the case of two interacting sites. Future efforts will focus on obtaining a relatively accurate estimate of the interaction parameters of the full model proposed in the text. This will permit us to simulate the kinetics of polymerization for the model, including the relevant nucleation mechanisms, and to compare these results with known experimental observations of sickle hemoglobin. As noted in the introduction, by understanding the conditions on which the nucleation depends, one might then be able to slow down the nucleation rate such as to prevent polymerization from occurring while HbS is in its deoxygenated state.

## ACKNOWLEDGMENTS

This material is based upon work supported by the G. Harold Mathers and Leila Y. Mathers Foundation and by the National Science Foundation, Grant No. DMR-0302598.

- <sup>1</sup>F. A. Ferrone, J. Hofrichter, and W. A. Eaton, *J. Mol. Biol.* **183**, 611 (1985).
- <sup>2</sup>G. Agarwal, J. C. Wang, S. Kwong, F. A. Ferrone, R. Josephs, and R. W. Briehl, *J. Mol. Biol.* **322**, 395 (2002).
- <sup>3</sup>O. Galkin and P. G. Vekilov, *J. Mol. Biol.* **336**, 43 (2004).
- <sup>4</sup>P. D. Ross, J. Hofrichter, and W. A. Eaton, *J. Mol. Biol.* **115**, 111 (1977).
- <sup>5</sup>W. A. Eaton and J. Hofrichter, *Adv. Protein Chem.* **40**, 63 (1990).
- <sup>6</sup>P. L. San Biagio and M. U. Palma, *Biophys. J.* **60**, 508 (1991).
- <sup>7</sup>A. Shirayev, D. L. Pagan, J. D. Gunton, D. S. Rhen, A. Saxena, and T. Lookman, *J. Chem. Phys.* **122**, 234911 (2005).
- <sup>8</sup>P. R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, *J. Chem. Phys.* **104**, 9932 (1996).
- <sup>9</sup>P. R. ten Wolde and D. Frenkel, *Science* **277**, 1975 (1997).
- <sup>10</sup>P. R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, *J. Chem. Phys.* **110**, 1591 (1999).
- <sup>11</sup>M. S. Wertheim, *J. Chem. Phys.* **87**, 7323 (1987).

- <sup>12</sup>G. Jackson, W. G. Chapman, and K. E. Gubbins, *Mol. Phys.* **65**, 1 (1988).
- <sup>13</sup>N. Asherie, A. Lomakin, and G. B. Benedek, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9465 (1999).
- <sup>14</sup>R. P. Sear, *Phys. Rev. E* **59**, 6838 (1999).
- <sup>15</sup>R. A. Curtis, H. W. Blanch, and J. M. Prausnitz, *J. Phys. Chem. B* **105**, 2445 (2001).
- <sup>16</sup>N. Kern and D. Frenkel, *J. Chem. Phys.* **111**, 9882 (2003).
- <sup>17</sup>E. A. Padlan and W. E. Love, *J. Biol. Chem.* **260**, 8280 (1985).
- <sup>18</sup>D. J. Harrington, K. Adachi, and W. E. Royer, Jr., *J. Mol. Biol.* **272**, 398 (1997).
- <sup>19</sup>J. S. van Duijneveldt and D. Frenkel, *J. Chem. Phys.* **4655**, 4655 (1992).
- <sup>20</sup>To make the free energy profile smoother it would be necessary to simulate more particles. ten Wolde and Frenkel (Ref. 9) used about  $10^4$  particles in their simulations. Anisotropic interaction make the simulations more computationally time consuming; thus we used a smaller number of particles.
- <sup>21</sup>C. Vega and P. A. Monson, *J. Chem. Phys.* **109**, 9938 (1998).