PHYSICAL REVIEW E **88**, 012721 (2013)

Crystallization of asymmetric patchy models for globular proteins in solution

Diana Fusco[1] and Patrick Charbonneau[1,2,3,*]
[1]*Program in Computational Biology and Bioinformatics, Duke University, Durham, North Carolina 27708, USA*
[2]*Department of Chemistry, Duke University, Durham, North Carolina 27708, USA*
[3]*Department of Physics, Duke University, Durham, North Carolina 27708, USA*

(Received 2 January 2013; revised manuscript received 21 June 2013; published 18 July 2013)

Asymmetric patchy particle models have recently been shown to describe the crystallization of small globular proteins with near-quantitative accuracy. Here, we investigate how asymmetry in patch geometry and bond energy generally impacts the phase diagram and nucleation dynamics of this family of soft matter models. We find the role of the geometry asymmetry to be weak, but the energy asymmetry to markedly interfere with the crystallization thermodynamics and kinetics. These results provide a rationale for the success and occasional failure of the proposal of George and Wilson for protein crystallization conditions as well as physical guidance for developing more effective protein crystallization strategies.

DOI: 10.1103/PhysRevE.88.012721 PACS number(s): 87.15.ak, 82.70.Dd, 87.15.nt

I. INTRODUCTION

Proteins are key biological molecules whose physiological roles are, for the most part, tightly linked to their three-dimensional structure. Because x-ray and neutron crystallography are the most widely used techniques to detail these structures [1,2], the difficulty of obtaining diffraction-quality protein crystals severely limits our understanding of living systems. Crystallizing a protein typically involves placing a drop of protein solution near a high-salt aqueous buffer that drives the vapor diffusion away from the drop. The nonvolatile solutes then steadily concentrate and, if the initial conditions are properly chosen, a protein crystal assembles [2]. From a physical point of view, identifying successful crystallization conditions is thus equivalent to determining the protein's solution phase diagram. The limited usefulness of existing physical descriptions and of knowledge-based approaches [3,4], however, leaves a vast space of experimental conditions to be screened. A material understanding of protein assembly is thus essential to developing more effective crystallization strategies.

Soft matter descriptions of protein assembly based on particles with isotropic, short-range attractive interactions [5–7]—as suggested by early structural studies [8–10]—provide some conceptual guidance. They identify the region between the solubility line, above which the solution is stable, and the liquid-liquid critical point, well below which the system precipitates into amorphous materials [11,12], as the "crystallization gap" where crystal assembly is possible. This schematic picture is, however, unable to reproduce many experimental trends [13–17]. The introduction of bond directionality in symmetric "patchy" models is aimed at better representing the effective protein-protein interactions that drive their crystallization [18,19]. Yet the most commonly studied versions of these models have symmetric and interchangeable patches, which are atypical of real proteins [4,20–22] and insufficient to describe the assembly of even the simplest of globular proteins [22,23]. In this article, we investigate the role of patch geometry and bond energy asymmetry on the phase diagram and assembly dynamics of a coarse-grained protein model of rigid globular proteins in aqueous solution. This additional anisotropy "direction" complements earlier experimental proposals [24]. It also maps onto the assembly of more complex structures in systems such as DNA-coated colloidal particles, in which the strength of directional interaction can be finely tuned [25–27]. Our work, therefore, identifies general regions of parameter space that should be targeted for specific colloidal assemblies, such as gel and crystal formation.

The plan for the paper is as follows. In Sec. II, we describe the model used. In Sec. III, we analyze the phase diagrams of a collection of models. In Sec. IV, we study nucleation and the pathways to crystallization. Finally in Sec. V, we determine how percolation interferes with crystallization.

II. MODEL DESCRIPTION

We describe each protein as a hard sphere of diameter σ, which sets the unit of length, with interacting directional patches that mimic the effective interactions between solvated proteins. This schematic description assumes that proteins maintain their structure throughout crystallization and that solvated electrolytes screen long-range electrostatic interactions, which is typical of protein solutions that produce diffraction quality crystals [28,29]. Crystallization cocktails that include salt as only cosolute indeed account for nearly 50% of successful experimental conditions in typical databases [30]. In these conditions, attraction is triggered by the specific chemical details at the protein surface and thus directional interactions dominate. This treatment complements and supports previous studies that focused on the interplay between specific and nonspecific (depletion) interactions [31–33].

We consider a variant of the patchy model of Ref. [34] in which patch-patch interactions are specific [35] and their range and strength are independently tunable [36]. The pairwise interaction between particles 1 and 2, whose centers are a distance r_{12} apart, is

$$\phi(r_{12},\Omega_1,\Omega_2) = \phi_{\mathrm{HS}}(r_{12}) + \sum_{i=1}^{n}[\phi_{2i,2i-1}(r_{12},\Omega_1,\Omega_2) + \phi_{2i-1,2i}(r_{12},\Omega_1,\Omega_2)], \quad (1)$$

---

[*] Corresponding author: patrick.charbonneau@duke.edu

where Ω_1 and Ω_2 are the Euler angles and n is the number of pairs of patches. A hard-sphere (HS) potential captures volume exclusion,

$$\phi_{\rm HS}(r) = \begin{cases} \infty, & r \leqslant \sigma, \\ 0, & r > \sigma. \end{cases} \quad (2)$$

The patch-patch interaction is the product of radial and angular components,

$$\phi_{2i,2i-1}(r_{12},\Omega_1,\Omega_2) = \psi_i(r_{12})\omega_{2i,2i-1}(\Omega_1,\Omega_2), \quad (3)$$

where

$$\psi_i(r) = \begin{cases} -\epsilon_i, & r \leqslant \lambda_i\sigma, \\ 0, & r > \lambda_i\sigma \end{cases} \quad (4)$$

and

$$\omega_{2i,2i-1}(\Omega_1,\Omega_2) = \begin{cases} 1, & \theta_{1,2i} \leqslant \delta_{2i} \quad \text{and} \quad \theta_{2,2i-1} \leqslant \delta_{2i-1}, \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

The interaction range λ_i is in units of σ, δ_{2i} is the semiwidth of patch $2i$, and $\theta_{1,2i}$ is the angle between the vector $\mathbf{r_{12}}$ and the vector defining patch $2i$ on particle 1. By symmetry, an analogous definition holds for $\theta_{2,2i-1}$. Here, the short radial extent of the square-well attraction, $\lambda_i = 1.1\sigma$ [37], and its surface coverage measured by the semiopening angle of its conical segment, $\delta_i = \cos^{-1}(0.89)$, are chosen to be typical of protein-protein interactions [20,23]. In contrast, the patch position on the surface and the bond energy ϵ_i are randomly chosen, under the sole constraint that the lattice formed by simply bonding the patches is the orthorhombic $P2_12_12_1$. This lattice, which is the most commonly observed in monomeric protein crystals [38,39], has three nonintersecting twofold screw axes that guarantee a high number of rigid-body degrees of freedom with minimal symmetry constraints.

We summarize the patch properties with energy and geometry asymmetry parameters,

$$\zeta = \frac{(\epsilon_1-\epsilon_2)^2+(\epsilon_1-\epsilon_3)^2+(\epsilon_2-\epsilon_3)^2}{2\left(\epsilon_1^2+\epsilon_2^2+\epsilon_3^2\right)},$$

$$\gamma = \frac{(I_1-I_2)^2+(I_1-I_3)^2+(I_2-I_3)^2}{2\left(I_1^2+I_2^2+I_3^2\right)},$$

where I_i represents the ith eigenvalue of the inertia tensor of the object represented in Fig. 1. Each patch (in red) carries a mass M at its center. The inertia tensor is computed over the set of weighted patches. The expression for γ guarantees that its value does not depend on the fictitious mass M nor on the radius of the particle, as long as they do not vary from one patch to the other [40]. Note that patches located on a perfect octahedron have $I_1 = I_2 = I_3$, and consequently $\gamma = 0$. Both ζ and $\gamma \in [0,1]$, where 0 corresponds to an equal energy distribution ($\epsilon_1 = \epsilon_2 = \epsilon_3$) and cubically distributed patches, and 1 corresponds to a complete energy asymmetry ($\epsilon_1 = \epsilon_2 = 0$ and $\epsilon_3 = \epsilon_{\rm tot}$) and a unit cell elongated in a single direction. Because of the $P2_12_12_1$ symmetry constraint on the crystal, patches cannot be too close to one another. Otherwise, bonded particles would overlap and the unit cell would stretch beyond the range of attraction $\lambda = 1.1\sigma$, which limits the achievable asymmetry and sets $\gamma \lesssim 0.1$. Because a cubic symmetry ($\gamma = 0$ and limiting case of $P22_12_1$) is not realizable within the three screw axes symmetry of $P2_12_12_1$, γ is limited from below as well. The adopted range and width of the interaction and the $P2_12_12_1$ symmetry ensure that two particles can only interact one bond at a time. This condition, together with the impossibility for a patch to interact with its copy, also prevents dimerization. Note that because the definition of γ and ζ is purely geometrical, there is no reason to expect that different models with identical asymmetry parameters should behave identically. Tables I and II summarize the parameter values used in this work.

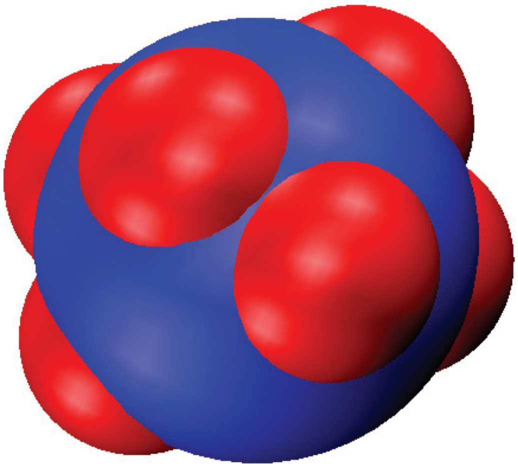

FIG. 1. (Color online) Sketch of a patchy particle. To determine the inertia tensor, we treat the patches as spherical balls [red (light gray)] whose center is at the particle surface.

III. PHASE DIAGRAM

For these 30 randomly selected sets of patch geometry and bond energies, we numerically determine the solubility line using free-energy integration and the metastable vapor-liquid line using Gibbs ensemble Monte Carlo simulations. We then compare the results to Wertheim's theory predictions.

A. Phase diagrams from simulations

Gibbs ensemble MC simulations (GEMC) directly determine the coexistence densities of the metastable gas and liquid phases [41]. We simulate a total of $N = 1000$ particles for 10^6 MC cycles, where each cycle consists on average of N particle displacements, N particle rotations, $N/10$ particle swaps, and two volume V changes. The critical temperature T_c and density are then estimated using the law of rectilinear diameters [42].

Because the gas-liquid line is metastable, crystallization happens so quickly for low-energy asymmetry that determining the gas and liquid coexistence densities is numerically inaccessible. In such cases, we estimate the critical temperature from Wertheim's perturbation theory (WPT) (see below).

To determine the fluid-solid coexistence curve, we integrate the Clausius-Clapeyron equation starting from one coexistence point using a fourth-order predictor-corrector algorithm [5].

TABLE I. Geometry parameters: the three numbers are the unit vector coordinates of the center of each patch. The center of patch 0 interacts with patch 1, patch 2 with patch 3, and patch 4 with patch 5. The case $\gamma = 0$ is reported for clarity.

γ	Patch$_0$	Patch$_1$	Patch$_2$	Patch$_3$	Patch$_4$	Patch$_5$
0	1.0	−1.0	0.0	0.0	0.0	0.0
	0.0	0.0	1.0	−1.0	0.0	0.0
	0.0	0.0	0.0	0.0	1.0	−1.0
0.0172	−0.8036	0.8036	−0.5186	−0.5186	0.3081	0.3081
	−0.5042	−0.5042	0.2731	0.2371	−0.8084	0.8084
	−0.3163	−0.3163	0.8102	−0.8102	0.5016	0.5016
0.0217	−0.7904	0.7904	−0.4227	−0.4227	0.3571	0.3571
	−0.5184	−0.5184	0.2807	0.2897	−0.776	0.776
	−0.3263	−0.3263	0.8617	−0.8617	0.5199	0.5911
0.0381	−0.7191	0.7191	−0.3813	−0.3813	0.3475	0.3475
	0.5659	0.5659	−0.1669	−0.1669	−0.7668	0.7668
	−0.4032	−0.4032	−0.9093	0.9093	0.5397	0.5397
0.0631	−0.6167	0.6167	0.6103	0.6103	−0.006	−0.006
	0.6521	0.6521	−0.3099	−0.3099	−0.9579	0.9579
	−0.441	−0.441	0.729	−0.729	0.2871	0.2871
0.0787	−0.9042	0.9042	−0.6276	−0.6276	0.2859	0.2859
	−0.3335	−0.3335	0.5386	0.5386	−0.9074	0.9074
	−0.2669	−0.2669	0.5622	−0.5662	0.3081	0.3081

The coexistence point itself is determined using free-energy calculations. The free energy of the fluid is computed using thermodynamic integration from the free energy of an ideal gas [43]. The free energy of the crystal is determined using an Einstein crystal with a fixed center of mass as in Ref. [44]. Its Hamiltonian is

$$H^{\mathrm{Ein}}(\Xi_{\mathrm{trans}},\Xi_{\mathrm{or}}) = \Xi_{\mathrm{trans}} \sum_{i=1}^{N} (\mathbf{r}_i - \mathbf{r}_{i,0})^2 + \Xi_{\mathrm{or}} \sum_{i=1}^{N} f(\theta_i,\phi_i,\chi_i),$$

where $f(\theta_i,\phi_i,\chi_i) = 1 - \cos(\psi_{i,1}) + 1 - \cos(\psi_{i,2})$, (θ_i,ϕ_i,χ_i) are the Euler angles describing the orientation of particle i, and $\psi_{i,j}$ is the angle formed between the vector defining patch j of particle i and the corresponding vector in the Einstein crystal. As explained in Ref. [43], the Helmholtz free energy of the reference Einstein crystal can then be written as

$$a_{\mathrm{Ein}}^{\mathrm{COM}} = a_{\mathrm{trans}}^{\mathrm{COM}} + a_{\mathrm{or}}^{\mathrm{COM}}, \tag{6}$$

TABLE II. Energy parameters with $\epsilon_{\mathrm{tot}} = 6$.

ζ	ϵ_1	ϵ_2	ϵ_3
0.00	2.0	2.0	2.0
0.11	1.2462	2.5482	2.2056
0.21	2.1	2.9066	0.9934
0.33	0.4854	2.8266	2.688
0.49	3.5756	0.2037	2.2207
0.50	3.0	3.0	0.0
0.55	3.96	1.8	0.24
0.64	4.32	1.5	0.18
0.79	4.8	0.6	0.6

where

$$\beta a_{\mathrm{trans}}^{\mathrm{COM}} = -\frac{3}{2}\frac{N-1}{N}\ln\left(\frac{\pi}{\beta\Xi_{\mathrm{trans}}}\right) - \frac{3}{2N}\ln N \tag{7}$$

and

$$\beta a_{\mathrm{or}}^{\mathrm{COM}} = -\ln\left\{\frac{1}{8\pi^2}\int d\theta\,\sin(\theta) d\phi\, d\chi\, \exp[-\beta\Xi_{\mathrm{or}} f(\theta,\phi,\chi)]\right\}.$$

The calculation of $a_{\mathrm{trans}}^{\mathrm{COM}}$ is straightforward, but that of $a_{\mathrm{or}}^{\mathrm{COM}}$ requires either a tedious numerical integration or an analytical approximation. We opt for the latter using a saddle-point approximation, which is accurate and efficient for the high values of $\beta\Xi_{\mathrm{or}}$ used here, because the integrand is then sharply peaked. Defining (θ_0,ϕ_0,χ_0) as the reference orientation in the Einstein crystal and changing the variable $\boldsymbol{\alpha} = (\cos(\theta),\phi,\chi)$ gives

$$\begin{aligned}\int d\theta\,\sin(\theta) d\phi\, d\chi\, \exp[-\beta\Xi_{\mathrm{or}} f(\boldsymbol{\alpha})] &= \int d\boldsymbol{\alpha}\, \exp[-\beta\Xi_{\mathrm{or}} f(\boldsymbol{\alpha})] \\ &\approx \frac{\exp[-\beta\Xi_{\mathrm{or}} f(\boldsymbol{\alpha}_0)](2\pi)^{3/2}}{(\beta\Xi_{\mathrm{or}})^{3/2}\det\{H[f(\boldsymbol{\alpha}_0)]\}^{1/2}} \\ &= \frac{(2\pi)^{3/2}}{(\beta\Xi_{\mathrm{or}})^{3/2}\det\{H[f(\boldsymbol{\alpha}_0)]\}^{1/2}},\end{aligned}$$

such that

$$\beta a_{\mathrm{or}}^{\mathrm{COM}} \approx \tfrac{3}{2}\ln(\beta\Xi_{\mathrm{or}}) + \tfrac{1}{2}\ln\left(8\pi\,\det\{H[f(\boldsymbol{\alpha}_0)]\}\right), \tag{8}$$

where $\det\{H[f(\boldsymbol{\alpha}_0)]\}$ is the determinant of the Hessian of function f computed at $\boldsymbol{\alpha}_0$. Its analytical expression is reported in Appendix A. Once the free energy of the reference crystal is known, the free energy of the actual crystal is obtained following a standard free-energy integration protocol [45].

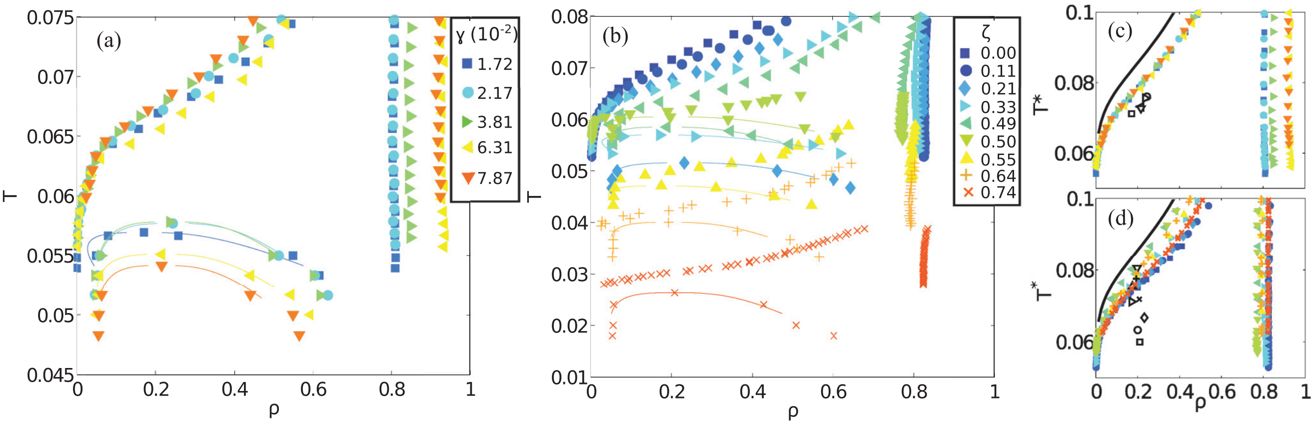

FIG. 2. (Color online) Temperature-density ρ phase diagrams of patchy models with (a) different γ at fixed $\zeta = 0.33$, and (b) with different ζ at fixed $\gamma = 0.0172$. (c) and (d) depict the same phase diagrams with T rescaled following WPT. The crystal-fluid coexistence lines (symbols) are then close to, yet distinct from, the WPT solubility line (solid black line). For visual clarity, only the gas-liquid critical points (black symbols) are reported in (c) and (d). Spontaneous crystallization during the simulations prevents the precise evaluation of the gas-liquid line for low ζ. The specific parameter values are given in Tables I and II.

Several simulations along an isobar starting from the fluid and from the crystal are then necessary to determine the temperature at which the chemical potential of the two phases coincides [43,45].

Figure 2 illustrates the simulated phase diagrams. The gas-liquid critical temperature T_c generally decreases with increasing γ because patch proximity anticorrelates bond formation and decreases the liquid entropy, although the limited number of systems studied partially hides this feature (Table III). The solubility line, by contrast, is clearly similar for different geometries at fixed ζ and monotonically shifts to lower temperatures with increasing energy asymmetry.

B. Phase diagrams from Wertheim's perturbation theory

According to Wertheim's perturbation theory [46,47], the fluid free energy can be approximated by the hard-sphere free energy plus a bond free-energy correction,

$$a_f = a_{\mathrm{HS}} + a_{\mathrm{bond}}, \tag{9}$$

TABLE III. Critical temperatures T_c for the models studied. An asterisk indicates that the system crystallized spontaneously in GEMC simulations and Wertheim's estimate is reported instead. No value indicates models for which the phase diagram was not determined. Temperatures are in units of ϵ_{tot}.

ζ \ γ	0.0172	0.0217	0.0381	0.0631	0.0787
0.00*	0.052	0.052	0.052	0.052	0.052
0.11*	0.053	0.053	0.053	0.053	0.053
0.21	0.052				
0.33	0.057	0.058	0.058	0.055	0.054
0.49	0.059	0.057	0.053	0.050	0.055
0.50	0.061				
0.55	0.047				
0.64	0.040				
0.79	0.026	0.024	0.024	0.019	0.020

where

$$\beta a_{\mathrm{bond}} = \sum_{a\in\Gamma}\left(\ln X_a - \frac{X_a}{2}\right) + \frac{m}{2}. \tag{10}$$

Here m is the total number of attractive sites, X_a is the probability that the molecule is not bonded at site a, and Γ is the set of interacting patches.

Similarly, the chemical potential is given by

$$\beta\mu_f = \beta a_f + \frac{\beta p}{\rho} = \beta a_{\mathrm{HS}} + \beta a_{\mathrm{bond}} + \frac{\beta p_{\mathrm{HS}}}{\rho} + \frac{\beta p_{\mathrm{bond}}}{\rho}, \tag{11}$$

where the pressure p contribution to bonding is

$$\beta p_{\mathrm{bond}} = \rho^2 \sum_{a\in\Gamma}\left(\frac{\partial X_a}{\partial \rho}\right)\left(\frac{1}{X_a} - \frac{1}{2}\right). \tag{12}$$

In the solid, $\beta a_s \approx \beta\mu_s$, because the ratio $\frac{\beta p_s}{\rho_s}$ is small [35,48]. The energetic contribution to the free energy is the energy of the fully bonded system $-\beta\epsilon_{\mathrm{tot}}$, while the entropic term is approximated using the range of interaction and the width of the patches [35],

$$\beta\mu_s = \beta a_s = -3\ln(\lambda - 1) - \ln\left(\frac{\delta^3}{\pi^2}\right) - \beta\epsilon_{\mathrm{tot}}. \tag{13}$$

At phase coexistence, the temperature, pressure, and chemical potential of the fluid and solid phases have to be identical. The pressure of the solid is once again ignored, so the only remaining constraint is $\beta_{\mathrm{coex}}\mu_f = \beta_{\mathrm{coex}}\mu_s$. Using the equations above, it follows that

$$\beta_{\mathrm{coex}} a_{\mathrm{HS}} + \beta a_{\mathrm{bond}} + \frac{\beta_{\mathrm{coex}} p_{\mathrm{HS}}}{\rho} + \frac{\beta_{\mathrm{coex}} p_{\mathrm{bond}}}{\rho}$$
$$= -3\ln(\lambda - 1) - \ln\left(\frac{\delta^3}{\pi^2}\right) - \beta_{\mathrm{coex}}\epsilon_{\mathrm{tot}}. \tag{14}$$

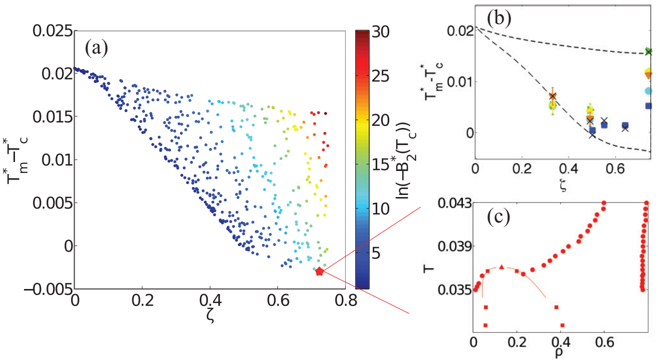

FIG. 3. (Color online) (a) WPT difference between rescaled melting temperature T_m^* and rescaled critical temperature T_c^* at the critical density $\rho_c \approx 0.2$, i.e., the crystallization gap. Each circle represents a distinct $\{\epsilon_i\}$ realization and is colored by its $\ln[-B_2(T_c)]$ value (higher values in the top right corner). (b) WPT predictions (black crosses) compared with simulation results for different patch geometries [symbols as in Fig. 2(a)]. The area within the dashed lines indicates the values covered by WPT in (a). (c) Phase diagram for $\{\epsilon_1 = 4.6655, \epsilon_2 = 1.2908, \epsilon_3 = 0.0437\}$ for which WPT predicts a stable gas-liquid coexistence. Even for an interaction range and patch coverage that would normally result in a metastable gas-liquid line, the bond energy asymmetry can lift T_c above the solubility line (Appendix B).

As the hard-sphere system itself is temperature-independent, it holds that

$$\beta_{\text{coex}}\left(a_{\text{bond}} + \frac{p_{\text{bond}}}{\rho} + \epsilon_{\text{tot}}\right) = C(\rho), \tag{15}$$

where

$$C(\rho) = -\beta a_{\text{HS}} - \frac{\beta p_{\text{HS}}}{\rho} - 3\ln(\lambda - 1) - \ln\left(\frac{\delta^3}{\pi^2}\right) \tag{16}$$

is a function that only depends on ρ. It thus follows that $a_{\text{bond}} + \frac{p_{\text{bond}}}{\rho} + \epsilon_{\text{tot}}$ represents a good temperature rescaling factor to obtain the master solubility line across the different models.

When compared with the simulated results, WPT overestimates the solubility temperature at all densities ρ, but nonetheless remarkably collapses the simulation results [Figs. 2(c) and 2(d)]. The numerical validation of WPT's T_c predictions—accurate to within 10–15%—allows us to estimate the size of the "crystallization gap" for a broader variety of models (Fig. 3) [49]. Interestingly, we find that for patch energy sets $\{\epsilon_i\}$ giving the same ζ, a lower second virial coefficient B_2 results in a larger crystallization gap (Fig. 3). In contrast to the proposal by George and Wilson (GW) that $\ln(-B_2^*) < 5$ identifies facile crystallization [3], the asymmetric models reveal that B_2 does not by itself set the size of the crystallization gap. The proposal is thus reasonable at low ζ, but breaks down at high ζ, where it even encompasses systems for which the critical point is fully stable [red star in Figs. 3(a) and 3(c), Appendix B]. In such systems, access to the crystal from a slowly concentrating, low-density solution would have to sidestep the metastable gas-liquid coexistence regime. This regime typically prevents the formation of all but the smallest crystallites [50,51]. High interaction asymmetry therefore provides a microscopic rationale for the failure of the GW proposal [52,53], which complements and supports previous suggestions that were based on a balance of specific and nonspecific interactions [31,32].

IV. CRYSTALLIZATION

Even if crystallization is thermodynamically possible, the free energy drive may be insufficient to induce a phase transition on experimentally relevant time scales. The role of asymmetry on homogeneous nucleation is thus examined.

We consider systems near their critical density $\rho_c \approx 0.2$ at different degrees of supersaturation $\eta = \frac{T_m - T}{T_m - T_c}$, where T_m is the solubility temperature at that density. We determine the size of the crystal clusters in the simulation box following a standard procedure that defines a *crystal-like bond* and *crystal-like particles* [54]. Due to the highly specific patch-patch interactions of our model, we generally define a crystal-like bond between particles 1 and 2 when they are actually bonded: $r < \lambda\sigma$, $\theta_{1,2i} < \delta$, and $\theta_{2,2i-1} < \delta$ for some i. A particle is considered to be crystal-like if it has six crystal-like bonds, and two crystal-like particles belong to the same crystal cluster if a crystal-like bond connects them. Visual inspection of these "crystals" confirms that the criterion selects actual crystal clusters. For the umbrella sampling simulations, we use a biasing harmonic potential with spring constant κ that varies between 0.07 and 0.12, depending on the model and the temperature studied,

$$H^{\text{bias}} = \kappa(s - s_0)^2, \tag{17}$$

where s is the size of the largest crystal cluster and s_0 is the target cluster size in the sampling window. Sampling windows are typically positioned every three particles, but denser sampling is sometimes required. The results of each simulation are then analyzed following a standard umbrella sampling protocol [54].

Unsurprisingly, the lower supersaturations correspond to higher free-energy barriers and larger critical nuclei [Fig. 4(c)]. Across various patch geometries, qualitatively similar results are obtained, but increasing the energy asymmetry significantly lowers the chemical potential difference, $\beta\Delta\mu$, between the fluid and the crystal. At high bond energy asymmetry, fewer patches dominate the energy of the two phases, which makes that contribution in the two phases more similar and reduces the drive to crystallize. Higher densities are then needed to obtain a comparable nucleation barrier. Although this effect is not a fundamental limitation for particles to crystallize, real proteins in high-density solutions may partially unfold and aggregate, which interferes with their crystallization [50]. In addition, at high ζ the narrow crystallization gap results in larger free-energy nucleation barriers. High-energy asymmetry thus hinders nucleation kinetics.

Classical nucleation theory (CNT) describes crystal formation reasonably well far above the critical point, but near and below T_c the assembly behavior is more complex. Previous studies of isotropically attractive systems have shown that well below T_c, spinodal decomposition leads to dynamical arrest [11,12], because spontaneous density fluctuations result in dense regions within which binding is irreversible. In similar systems near the critical point, "two-step" nucleation is favored [7]. Crystal formation is then easier in high-density than in low-density fluid regions. The corresponding assembly

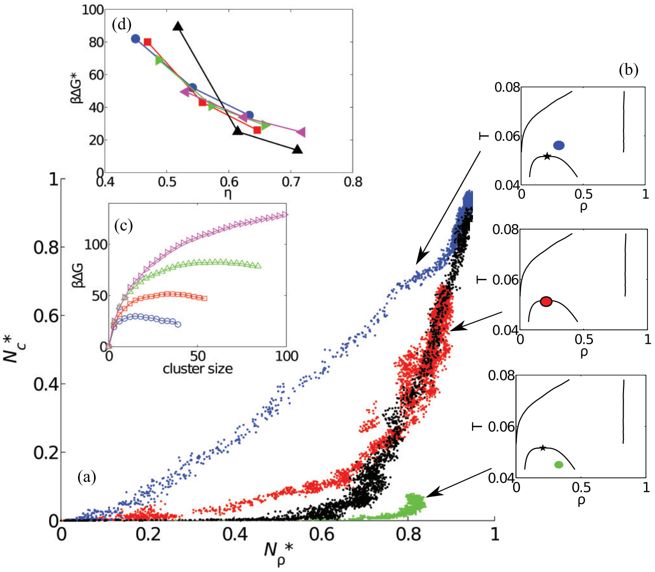

FIG. 4. (Color online) (a) Rescaled size of the largest liquid cluster N_ρ^* compared to that of the largest crystal cluster N_c^* for a model with $\zeta = 0$ and $\gamma = 0.017$ [blue (upper), red (dark gray), and green (light gray)] and $\zeta = 0.2$ and $\gamma = 0.017$ (black). The liquid cluster size is rescaled by the number of particles in the system ($N = 864$) and the crystal cluster size by the size of the largest crystal cluster that fits in the simulation box (600). The different trends correspond to initial homogeneous fluid configurations under different conditions, as illustrated in (b). (c) Above the critical point, nucleation barriers can be computed (blue circle $\eta = 0.37$, red squares $\eta = 0.46$, green triangles $\eta = 0.55$, and magenta right-pointed triangles $\eta = 0.64$). (d) The height of the nucleation barrier for different models (blue circles $\zeta = 0$ and $\gamma = 0.017$, red squares $\zeta = 0$ and $\gamma = 0.021$, green right-pointed triangles $\zeta = 0$ and $\gamma = 0.065$, black triangles $\zeta = 0.1$ and $\gamma = 0.017$, magenta left-pointed triangles $\zeta = 0.2$ and $\gamma = 0.017$).

behavior of patchy systems, whose low-density crystals may not be favored by spontaneous fluid density fluctuations [31], is studied here in unbiased constant NpT MC simulations. These simulations sketch out the minimum free-energy path for the assembly, which we track along the largest drop and the largest crystal cluster reaction coordinates [Fig. 4(a)]. The largest crystal cluster is determined as described above. Similarly, *liquid-like particles* are defined as those that have at least four close neighbors (particles whose centers are within $\lambda\sigma$ of each other). Two *liquid-like particles* belong to the same *liquid cluster* if they are close neighbors. These trajectories follow a fictitious dynamics without accounting for collective moves and where time should be properly rescaled. As previously shown [55], such trajectories are representative of Brownian dynamics configurational space sampling for sufficiently short steps. The robustness of our observations is also confirmed by repeating the simulations using the virtual-move MC of Ref. [56], which allows for collective rearrangements (for details, see Appendix C) [31–33]. We obtain for the symmetric case, $\zeta = 0$, far above T_c, that the largest cluster formation is always crystalline and CNT applies. Near the critical point (within ~10% of T_c), a growing liquid drop first forms and only subsequent structural reorganization of the many microcrystals results in a large crystal cluster.

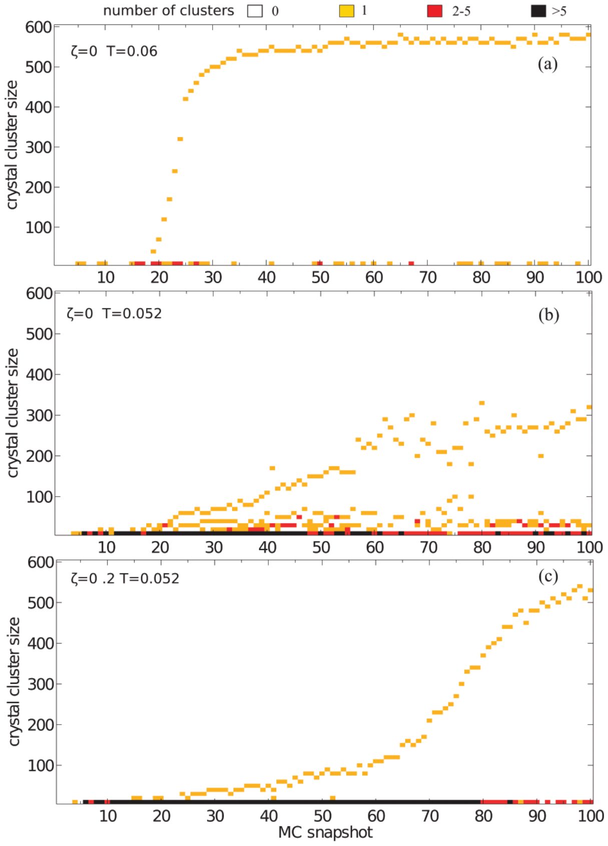

FIG. 5. (Color online) Distribution of crystal cluster sizes along the unbiased NpT MC simulations, respectively, for (a) $\zeta = 0$ at $T > T_c$, (b) $\zeta = 0$ at $T = T_c$, and (c) $\zeta = 0.2$ at $T = T_c$ for a system of 864 particles. The distribution of crystal cluster size from instantaneous snapshots is represented. White indicates the lack of clusters of that size, orange (lighter gray) indicates one cluster being present, red (darker gray) indicates between two and five clusters (few), and black indicates that more than five clusters are present (many). Five snapshots cover 10^5 MC sweeps in the first two panels, and 5×10^5 in the last panel.

We can gain additional physical insights into the dynamical assembly pathway by characterizing the distribution of crystal clusters within the largest liquid drop. Figure 5 reports the distribution of crystal cluster sizes from simulation snapshots. Panel (a) shows a classical nucleation scenario in which, after a waiting time (2×10^6 MC sweeps), a critical nucleus appears and grows rapidly. No secondary nucleation event is observed. Panel (b) illustrates the status of the system with symmetric interactions at the critical point. Almost instantaneously, microcrystals (with fewer than 50 particles per cluster) form, and many of them survive the whole simulation. The formation of the largest cluster is much less smooth than in the classical nucleation limit. Note that the size of the largest cluster does not grow monotonically. Noise results from the slow annealing of defects, which contrasts with what happens for a single nucleation event.

Figure 4 shows that the behavior at the critical point between low- and intermediate-energy asymmetry models is similar (black and red dots). Yet a closer analysis reveals that the cluster distribution exhibits a significant difference [Fig. 5(c)]. In the asymmetric case, a single nucleation event is followed by the growth of a single crystal cluster rather than the re-organization of many microcrystals. Despite this resemblance to classical nucleation, the time between the appearance of a first critical nucleus and its full growth is long compared to a nucleation scenario in which crystallization is rare yet rapid. Crystallization occurs on a time scale similar to percolation, and it is possible that the interplay between the two phenomena underlies the observed slow growth. It is interesting to note that we do not observe any crystal cluster of significant size above the critical temperature within the simulation time, even though the crystallization barrier height is similar to that of the symmetric case [Fig. 4(d)]. This feature is left for future enquiries.

V. PERCOLATION

We finally consider whether direct percolation dynamically competes with crystallization. Below the percolation threshold $T_p(\rho)$, the system forms infinitely large spanning networks that can be long-lasting when bonding is strong [57,58].

To explore the interplay between percolation and bond energy asymmetry (patch geometry asymmetry has only a weak impact), we determine $T_p(\rho_c)$ using finite-size rescaling [59]. We run 20 NVT simulations with, respectively, $N = 2048$, 4000, and 6912 for several temperatures at $\rho = 0.2$. During the simulation, we determine the size of the biggest network defined as the largest set of particles connected by at least two bonds. If such a cluster spans the whole simulation box along one dimension within 10^5 MC sweeps, the system is deemed percolating. The percolation probability is the fraction of simulations showing such a percolating cluster. The data are in agreement with the tabulated 3D critical exponent to within 1% [60,61]. For finite-size rescaling, we use the tabulated critical exponents and the standard procedure [59].

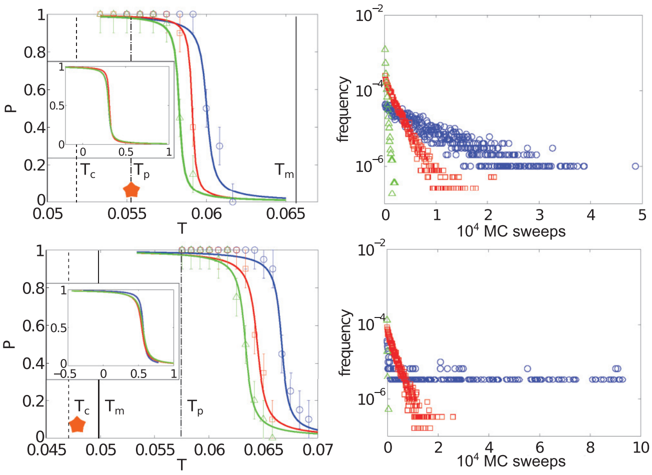

FIG. 6. (Color online) Percolation behavior for $\zeta = 0.2$ (top) and $\zeta = 0.55$ (bottom). Left panels show the probability P of observing a spanning network as a function of temperature for a system of size $N = 2048$ (blue circles), 4000 (red squares), and 6912 (green triangles). The superimposed vertical lines indicate the critical temperature (dashed), the melting temperature (solid), and the percolation temperature (dot-dashed) estimated by finite-size scaling (inset). The right panels show the distribution of bond lifetime in the network at $T = 0.055$ and 0.048, respectively (orange stars in the left panels). Blue circles indicate the strongest bond, red squares the intermediate bond, and green triangles the weakest bond.

Figure 6 shows the results for systems with relatively low ($\zeta = 0.22$) and high ($\zeta = 0.55$) asymmetry. In the first system, the percolation threshold lies just above T_c, while in the second, in which the strongest bond is much longer-lasting, T_p is well above the solubility line. The dynamical relevance of percolation on crystallization is estimated from the distribution of bond lifetimes within the crystallization gap. At low bond energy asymmetry, the rearrangement of all bonds is observed within a few thousand MC steps. At high asymmetry, in contrast, the lifetime of the strongest bond (blue circles) is comparable to the length of the simulation (10^5 MC sweeps). The network is frozen, the bonds are almost irreversible, and no rearrangement takes place. This observation suggests that identifying the crystallization gap may be insufficient for crystallizing particles with high-energy asymmetry, because a long-lasting gel caused by direct percolation then dynamically interferes with crystallization within the gap. Weakening the strongest bonds may be the only way to allow crystallization in these systems.

VI. CONCLUSION

To gain insights into protein crystallization and soft matter assembly more generally, we have considered the role of patch geometry and bond energy asymmetry on the crystal assembly of a family of schematic models. We find patch geometry asymmetry to have a weak effect, but bond energy asymmetry to severely impede crystallization thermodynamics and kinetics. The crystallization gap shrinks, gel formation is favored, and nucleation shifts to higher supersaturations. The union of these observations suggests that to facilitate locating proper crystallization conditions, it is sometimes more effective to symmetrize the directional pair interactions between colloids or proteins, rather than specifically strengthen one of them, as is sometimes implicitly suggested [4,23]. It also offers a rationalization of the GW crystallization slot proposal as well as of its occasional failure. At low bond energy asymmetry, the B_2 prescription falls within the crystallization slot; at high asymmetry, a large crystallization gap is only observed for values of B_2 below the prescribed ones, which corresponds to long-living gels, while for B_2 within the slot the crystallization gap is very small or even negative. The GW crystallization prescription is therefore a necessary but insufficient condition for detecting optimal experimental conditions.

Although we are now markedly closer to understanding simple, monomeric protein crystallization, the assembly features of more complex proteins remain a challenge. Some proteins dimerize or change conformation on a time scale comparable to their crystallization, while membrane proteins typically require entirely different crystallization approaches from the one considered in this work. Further modifications to patchy particle models, such as self-interacting or dynamically evolving patches, may thus guide our understanding of the assembly of these complex yet crucial molecules.

ACKNOWLEDGMENT

P.C. acknowledges NSF support under Grant No. NSF DMR-1055586.

APPENDIX A: SADDLE-POINT APPROXIMATION

The analytical expression for $f(\boldsymbol{\alpha})$ in Eq. (8) is

$$
\begin{aligned}
f(\boldsymbol{\alpha}) = \frac{1}{2}\Big\{&4 - 3\sqrt{1-y^2}\cos(\phi)\cos(\phi_0)\sin(\theta_0) - \sqrt{1-y^2}\cos(\chi)\cos(\chi_0)\sin(\theta_0) + \sqrt{1-y^2}\cos(\chi)\cos(\chi_0)\cos^2(\zeta)\sin(\theta_0)\\
&-\sqrt{1-y^2}\cos(\phi)\cos(\phi_0)\cos(2\zeta)\sin(\theta_0) - 2\sqrt{1-y^2}\cos(\phi)\cos(\zeta)\sin(\phi_0)\sin(\chi_0)\sin(\zeta)\\
&-2y\cos(\chi_0)\cos(\zeta)\sin(\theta_0)\sin(\zeta) + 2y\cos(\phi)\cos(\chi)\sin(\phi_0)\sin(\chi_0)\sin^2(\zeta) - 2\cos(\phi)\cos(\phi_0)\sin(\chi)\sin(\chi_0)\sin^2(\zeta)\\
&-\sqrt{1-y^2}\cos(\chi)\cos(\chi_0)\sin(\theta_0)\sin^2(\zeta) + y\cos(\phi)\cos(\phi_0)\cos(\chi)\sin(\theta_0)\sin(2\zeta) + \cos(\phi)\sin(\phi_0)\sin(\chi)\sin(\theta_0)\sin(2\zeta)\\
&+\cos(\theta_0)\big[-3y + \cos(\phi_0)\cos(\chi_0)\sin(\phi)\sin(\chi) + \cos^2(\zeta)(-y + y\cos(\chi)\cos(\chi_0)\sin(\phi)\sin(\phi_0)\\
&-\cos(\phi_0)\cos(\chi_0)\sin(\phi)\sin(\chi)) - 2\sqrt{1-y^2}\cos(\chi)\cos(\zeta)\sin(\zeta) + y\sin^2(\zeta) + \cos(\phi_0)\cos(\chi_0)\sin(\phi)\sin(\chi)\sin^2(\zeta)\\
&-2\cos(\phi)\cos(\chi_0)\sin(\phi_0)\sin(\chi)\sin^2(\zeta) - y\cos(\chi)\cos(\chi_0)\big(2\cos(\phi)\cos(\phi_0)\sin^2(\zeta) + \sin(\phi)\sin(\phi_0)(1+\sin^2(\zeta))\big)\\
&+\sqrt{1-y^2}\cos(\phi)\cos(\phi_0)\cos(\chi_0)\sin(2\zeta) + \sqrt{1-y^2}\cos(\chi_0)\sin(\phi)\sin(\phi_0)\sin(2\zeta)\big]\\
&+\sin(\phi)\big[-\sin(\phi_0)\big(2\sin(\chi)\sin(\chi_0)\sin^2(\zeta) + \sin(\theta_0)\big(3\sqrt{1-y^2} + \sqrt{1-y^2}\cos(2\zeta) - 2y\cos(\chi)\cos(\zeta)\sin(\zeta)\big)\big)\\
&+\cos(\phi_0)(-2\cos(\zeta)\sin(\chi)\sin(\theta_0)\sin(\zeta) + \sin(\chi_0)(-2y\cos(\chi)\sin^2(\zeta) + \sqrt{1-y^2}\sin(2\zeta)))\big]\Big\}.
\end{aligned}
$$

APPENDIX B: GRAND-CANONICAL MC APPROXIMATION FOR THE CRITICAL POINT

To check the position of the critical point for the extreme case of Fig. 3(c) ($\epsilon_1 = 4.6655, \epsilon_2 = 1.2908$, and $\epsilon_3 = 0.0437$), we perform grand-canonical MC (GCMC) simulations. Because one of the patches is markedly stronger than the others, percolation takes place at a relatively high temperature. Consequently, GCMC samples phase space slowly and poorly, and thus the critical temperature estimate is affected by large errors. To obtain a better estimate of the phase diagram, we perform GCMC simulations for systems with an increasing strength of the strongest patch (keeping the other patches identical) and we fit a power law to the resulting critical temperatures (Fig. 7). The value of the fit for $\epsilon_1 = 4.6655$ ($T = 0.045$ in units of ϵ_{tot}) confirms the stability of the critical point with respect to the solubility line.

APPENDIX C: VIRTUAL MOVE MC

Standard MC simulations are based on sequential perturbation of the system and do not directly account for the collective moves through which the system sometimes relaxes. Although it has been shown that for small enough displacements, MC recovers the Brownian dynamics of patchy particle models [55], it is reasonable to wonder if collective moves could nonetheless affect the system's dynamics. To check this possibility, we implement the virtual move MC algorithm [56], which accommodates cluster displacement and rotations and prevents the system from becoming stuck in unphysical traps. This commonly used algorithm has been shown to reproduce real dynamics of short-range attractive systems and it is commonly used for this purpose [31–33,56].

A virtual move consists of identifying a cluster to randomly displace or rotate. Each displacement draws from a uniform distribution between 0 and 0.2σ, and each rotation uniformly selects an axis of rotation and an angle of rotation. Following Ref. [56], to avoid generating large clusters whose moves will often be rejected, we draw the cutoff n_c of the cluster size from

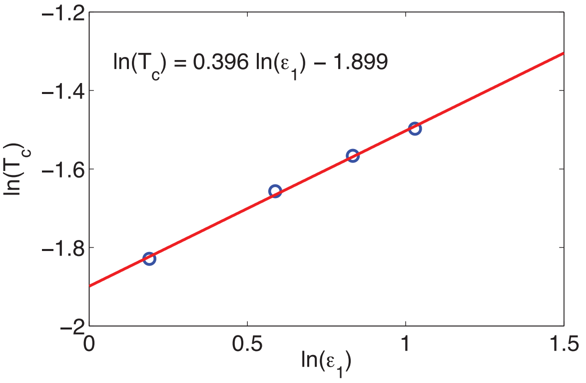

FIG. 7. (Color online) Fit of the critical point determined with GCMC simulations for systems with an increasingly strong strongest patch.

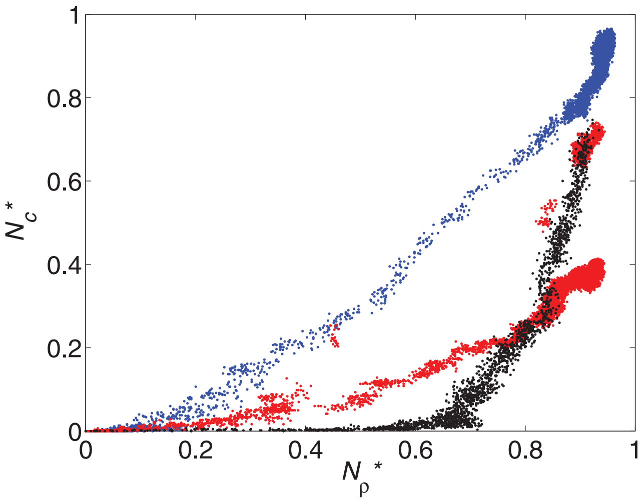

FIG. 8. (Color online) Crystallization pathways as in Fig. 3(a) using virtual move MC to simulate the dynamics. Comparison with Fig. 3(a) indicates that the crystallization pathways do not depend on the microscopic dynamics.

$P(n_c) \propto n_c^{-1}$. The results in Fig. 8 are in agreement with those generated by the standard NpT MC (Fig. 4), and they confirm the robustness of the phenomenology with respect to changes in the microscopic dynamics.

[1] N. E. Chayen, in *Advances in Protein Chemistry and Structural Biology*, edited by J. Andrzej, Vol. 77 (Academic, London, 2009) pp. 1–22.

[2] A. McPherson, *Crystallization of Biological Macromolecules* (CSHL Press, Cold Spring Harbor, NY, 1999).

[3] A. George and W. W. Wilson, Acta Crystallogr. D **50**, 361 (1994).

[4] Z. S. Derewenda, Structure **12**, 529 (2004).

[5] M. H. J. Hagen and D. Frenkel, J. Chem. Phys. **101**, 4093 (1994).

[6] D. Rosenbaum, P. C. Zamora, and C. F. Zukoski, Phys. Rev. Lett. **76**, 150 (1996).

[7] P. R. ten Wolde and D. Frenkel, Science **277**, 1975 (1997).

[8] J. Janin, Prog. Biophys. Mol. Biol. **64**, 145 (1995).

[9] O. Carugo and P. Argos, Protein Sci. **6**, 2261 (1997).

[10] R. P. Bahadur, P. Chakrabarti, F. Rodier, and J. Janin, J. Mol. Biol. **336**, 943 (2004).

[11] P. J. Lu, E. Zaccarelli, F. Ciulla, A. B. Schofield, F. Sciortino, and D. A. Weitz, Nature (London) **453**, 499 (2008).

[12] A. Fortini, E. Sanz, and M. Dijkstra, Phys. Rev. E **78**, 041402 (2008).

[13] C. Haas, J. Drenth, and W. W. Wilson, J. Phys. Chem. B **103**, 2808 (1999).

[14] A. Lomakin, N. Asherie, and G. B. Benedek, J. Chem. Phys. **104**, 1646 (1996).

[15] A. Lomakin, N. Asherie, and G. B. Benedek, Proc. Natl. Acad. Sci. USA **96**, 9465 (1999).

[16] R. A. Curtis, H. W. Blanch, and J. M. Prausnitz, J. Phys. Chem. B **105**, 2445 (2001).

[17] J. J. McManus, A. Lomakin, O. Ogun, A. Pande, M. Basan, J. Pande, and G. B. Benedek, Proc. Natl. Acad. Sci. USA **104**, 16856 (2007).

[18] C. Gögelein, G. Nägele, R. Tuinier, T. Gibaud, A. Stradner, and P. Schurtenberger, J. Chem. Phys. **129**, 085102 (2008).

[19] E. Bianchi, R. Blaak, and C. N. Likos, Phys. Chem. Chem. Phys. **13**, 6397 (2011).

[20] G. Pellicane, G. Smith, and L. Sarkisov, Phys. Rev. Lett. **101**, 248102 (2008).

[21] A. McPherson, *Introduction to Macromolecular Crystallography, Macromolecular Crystallography* (Wiley-Liss, Hoboken, NJ, 2003).

[22] N. Dorsaz, L. Filion, F. Smallenburg, and D. Frenkel, Faraday Disc. **159**, 9 (2012).

[23] D. Fusco, J. J. Headd, A. De Simone, and P. Charbonneau, arXiv:1206.6332 [cond-mat.soft] (unpublished).

[24] S. C. Glotzer and M. J. Solomon, Nat. Mater. **6**, 557 (2007).

[25] S. C. Glotzer, Science **306**, 419 (2004).

[26] F. J. Martinez-Veracoechea, B. M. Mladek, A. V. Tkachenko, and D. Frenkel, Phys. Rev. Lett. **107**, 045902 (2011).

[27] Y. Wang, D. R. Breed, V. N. Manoharan, L. Feng, A. D. Hollingsworth, M. Weck, and D. J. Pine, Nature (London) **491**, 51 (2012).

[28] G. E. Dale, C. Oefner, and A. DArcy, J. Struct. Biol. **142**, 88 (2003).

[29] L. Slabinski, L. Jaroszewski, A. P. Rodrigues, L. Rychlewski, I. A. Wilson, S. A. Lesley, and A. Godzik, Protein Sci. **16**, 2472 (2007).

[30] M. Charles, S. Veesler, and F. Bonneté, Acta Crystallogr. D **62**, 1311 (2006).

[31] T. K. Haxton and S. Whitelam, Soft Matter **8**, 3558 (2012).

[32] T. K. Haxton and S. Whitelam, Soft Matter **9**, 6851 (2013).

[33] S. Whitelam, Phys. Rev. Lett. **105**, 088102 (2010).

[34] J. Russo, J. M. Tavares, P. I. C. Teixeira, M. M. Telo da Gama, and F. Sciortino, Phys. Rev. Lett. **106**, 085703 (2011).

[35] R. P. Sear, J. Chem. Phys. **111**, 4800 (1999).

[36] N. Kern and D. Frenkel, J. Chem. Phys. **118**, 9882 (2003).

[37] M. G. Noro and D. Frenkel, J. Chem. Phys. **113**, 2941 (2000).

[38] S. W. Wukovitz and T. O. Yeates, Nat. Struct. Mol. Biol. **2**, 1062 (1995).

[39] The unit temperature T (and its inverse, β) is set by $\epsilon_1 + \epsilon_2 + \epsilon_3 = \epsilon_{\text{tot}}$ with Boltzmann's constant $k_B = 1$. The geometrical orientation varies from one pair of patches to the next. Although these models can crystallize in other lattices at high temperatures and densities, these structures are not relevant for protein crystallization.

[40] W. L. Miller and A. Cacciuto, J. Chem. Phys. **133**, 234903 (2010).

[41] A. Z. Panagiotopoulos, Mol. Phys. **61**, 813 (1987).

[42] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic, London, 2001).

[43] C. Vega, E. Sanz, J. L. F. Abascal, and E. G. Noya, J. Phys. Condens. Matter **20**, 153101 (2008).

[44] D. Frenkel and A. J. C. Ladd, J. Chem. Phys. **81**, 3188 (1984).

[45] F. Romano, E. Sanz, and F. Sciortino, J. Chem. Phys. **132**, 184501 (2010).

[46] M. S. Wertheim, J. Stat. Phys. **35**, 19 (1984).

[47] M. S. Wertheim, J. Stat. Phys. **35**, 35 (1984).

[48] G. Jackson, W. G. Chapman, and K. E. Gubbins, Mol. Phys. **65**, 1 (1988).

[49] We assume that the simulated model's critical density $\rho_c \approx 0.2$ is constant, which is numerically reasonable. Wertheim theory, however, notoriously underestimates the critical density for patchy particle models [62].

[50] P. G. Vekilov and A. A. Chernov, Solid State Phys. **57**, 1 (2002).

[51] N. Asherie, Methods **34**, 266 (2004).

[52] F. Bonnet, D. Vivars, C. Robert, and N. Colloch, J. Cryst. Growth **232**, 330 (2001).

[53] J. Blouwolff and S. Fraden, J. Cryst. Growth **303**, 546 (2007).

[54] I. Saika-Voivod, F. Romano, and F. Sciortino, J. Chem. Phys. **135**, 124506 (2011).

[55] C. De Michele, S. Gabrielli, P. Tartaglia, and F. Sciortino, J. Phys. Chem. B **110**, 8064 (2006).

[56] S. Whitelam and P. L. Geissler, J. Chem. Phys. **127**, 154101 (2007).

[57] J. M. Tavares, P. I. C. Teixeira, and M. M. Telo da Gama, Phys. Rev. E **81**, 010501 (2010).

[58] E. Zaccarelli, J. Phys.: Condens. Matter **19**, 323101 (2007).

[59] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor & Francis, London, 1992).

[60] C. D. Lorenz and R. M. Ziff, Phys. Rev. E **57**, 230 (1998).

[61] H. Neitsch and S. H. L. Klapp, J. Chem. Phys. **138**, 064904 (2013).

[62] E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli, and F. Sciortino, Phys. Rev. Lett. **97**, 168301 (2006).