

Supporting Information for “Control of pathways and yields of protein crystallization through the interplay of nonspecific and specific attractions”

Stephen Whitelam*

Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

Figs. S1, S2 and S3/S4/S5 supplement, respectively, Figs. 1, 2 and 3 (main text). Fig. S6 supplements the conclusions section of the main text.

Sampling simulations: Nonlocal ‘teleportation’ moves. The following algorithm was used to facilitate ‘sampling’ simulations by making moves that allow bond-making and bond-breaking to occur simultaneously, rather than sequentially, and so avoiding the long delay associated with the breaking of strong bonds. It also annuls to large extent the computational cost of diffusion. It is similar in spirit to the algorithm of Chen and Siepmann [1], though slightly different in detail.

Starting in microstate μ , we select one monomer, A, with uniform probability. We denote by n_μ the number of monomers in microstate μ that lie within A’s interaction range, regardless of the strength of their interaction with A. We then select with uniform probability a second monomer, B, distinct from A, and ‘teleport’ A to a position (chosen uniformly) lying within the interaction range of B. The identity of the recipient monomer B differs in general for forward and reverse moves. Upon teleportation we retain or randomize A’s orientation with equal likelihood. This move defines a proposed new microstate, ν .

We accept the proposed move of A in order to satisfy detailed balance. Detailed balance [2] guarantees convergence to the chosen stationary distribution by requiring that at thermal equilibrium the rate for passing from state μ to state ν is equal to the rate for the reverse transition:

$$\frac{e^{-\beta E_\mu}}{Z} p_{\text{gen}}(\mu \rightarrow \nu) p_{\text{acc}}(\mu \rightarrow \nu) = \frac{e^{-\beta E_\nu}}{Z} p_{\text{gen}}(\nu \rightarrow \mu) p_{\text{acc}}(\nu \rightarrow \mu). \quad (\text{S1})$$

Here E_μ is the energy of the system in state μ , and $Z \equiv \sum_\alpha e^{-\beta E_\alpha}$ is the partition function of the system. The variables p_{gen} and p_{acc} are the respective probabilities of generating and accepting a given transition. For the teleportation of A, the likelihood of generating the forward move $\mu \rightarrow \nu$ is, ignoring rotational degrees of freedom,

$$p_{\text{gen}}^{\text{teleport}}(\mu \rightarrow \nu) = \frac{1}{N} \frac{n_\nu}{N-1} \frac{1}{V_{\text{int}}}, \quad (\text{S2})$$

where V_{int} is the volume within which two particles may interact. In Equation (S2) the factor of N^{-1} is the probability of selecting particle A, while $n_\nu/(N-1)$ is the probability of choosing any particle that may bring A to its final state. The variable n_ν is equal to the number of neighbors of A in the *final* state. The factor of V_{int}^{-1} accounts for the probability of choosing a final position for A, given that a particle B has been chosen. The likelihood of generating the reverse move is

$$p_{\text{gen}}^{\text{teleport}}(\nu \rightarrow \mu) = \frac{1}{N} \frac{n_\mu}{N-1} \frac{1}{V_{\text{int}}}. \quad (\text{S3})$$

Inserting Equations (S2) and (S3) into Equation (S1) reveals the ratio of acceptance rates for forward and reverse moves required to enforce detailed balance. We choose to satisfy this ratio by accepting the proposed move of A with probability

$$p_{\text{acc}}^{\text{teleport}}(\mu \rightarrow \nu) = \min \left(1, \frac{n_\mu}{n_\nu} e^{-\beta \Delta E} \right), \quad (\text{S4})$$

where $\Delta E \equiv E_\nu - E_\mu$ is the energy change resulting from the nonlocal move of A. The ratio of the number of A's neighbors in each state, n_μ/n_ν , accounts for the fact that the number of particles that could have effected the teleport of A may differ for forward and reverse moves. Because A possesses at least one neighbor in the final state ν , the proposed move will be rejected if A possesses no neighbors in the initial state μ (isolated particles cannot move). Because of the latter restriction, the algorithm was combined with single-particle moves in order to ensure ergodic sampling.

Dynamic simulations. We drew monomer translation magnitudes from a uniform distribution with maximum $0.3a$, and rotations from a uniform distribution of 0.24 radians. We performed rotations and translations with equal likelihood. As a test, we performed for two sets of model parameters, $(\epsilon_n, \epsilon_d) = (2, 5)$ and $(0, 9)$, dynamic simulations in which the likelihood of translations, p_{trans} , ranged from 0.1 to 1. We consider the assembly behavior seen for our choice of $p_{\text{trans}} = 0.5$ to be representative of the behavior seen throughout most of that range. However, for p_{trans} greater than about 0.9, crystal yields dropped precipitously because particles failed to rotate sufficiently rapidly to find the crystal state.

Phase classifications and order parameters. We denote by f_c and f_p the fraction of crystalline and partially crystalline particles in a simulation box, i.e. those making respectively 3 and 2 directional bonds. Crystal order (squares) in Figure 1 (main text) is attributed to states with $\langle f_c \rangle > 0.1$. Because specific heat peaks associated with liquid-vapor critical fluctuations are relatively broad for the system sizes studied here, phase separation in Figure 1 (main text) is said to occur if the system is non-crystalline and if the thermal average of the fraction of particles residing in the system's largest cluster, $\langle f_{\text{max}} \rangle$, exceeds 0.5. The crossover point $\langle f_{\text{max}} \rangle = 0.5$ coincides approximately with the location of the relevant specific heat peak (see Figure S1). States displaying no crystalline order or evidence of phase separation are identified as homogeneous fluid. In Figure S1 we show that at large values of ϵ_n a non-crystalline tetratic phase emerges. We consider a tetratic order parameter $f_t \equiv (N(N-1))^{-1} \sum_{\langle ij \rangle} \cos(4\theta_{ij})$. Here the sum runs over all particle pairs in the largest cluster in the simulation box; θ_{ij} is the angle between the long axes of monomers i and j ; and N is the total number of particles in the system (normalization of f_t is arbitrary). The crystal possesses tetratic order and is described by a large value of f_c ; by contrast, the non-crystalline tetratic phase is described by a small value of f_c . In sampling simulations we used a non-crystalline tetratic phase as an initial condition, for moderate values of ϵ_n , in order to assess its stability.

In Figure 3 (main text) we report a scaled crystal yield $\hat{f}_c \equiv f_c (f_c / (f_p + f_c))^2$, designed to reward crystals possessing a large ratio of bulk-to-surface particles. Yields reported in terms of f_c are qualitatively similar, and give rise to similar conclusions (see e.g. Fig. S5). We report yields in Figure 3 (main text) for fixed processor times (200 CPU hours per simulation), because nucleation in mildly supercooled cases occurs only after a number of MC steps exceeding that accessible in a reasonable timescale within deeply supercooled, arrested simulations. However, we draw similar conclusions on the basis of yields reported after fixed numbers of MC steps (see Figure S3). Typically, some dynamic simulations in mildly supercooled regions produce equilibrated crystals, while independent simulations at the same thermodynamic state fail to give rise to a nucleation event.

We have not accounted for finite-size effects: crystal instability therefore means that the crystal's critical nucleus exceeds the number of particles in the simulation box. However, our aim is to probe mechanisms of assembly within simulations of finite size: 'sampling' simulations are designed to identify the largest possible yield of crystal accessible to such simulations (or to identify when phase separation in such systems might occur), and not to probe idealized bulk behavior.

Is the second virial coefficient a predictive measure of crystallization propensity? The second virial coefficient of a protein, B_2 , is a useful measure of its propensity for crystallization: many proteins that crystallize possess values of B_2 that lie within a well-defined range [3]. The second virial coefficient for our model S-layer system may be written,

up to constant terms, as $B_2 = -\kappa(A_n - A_d)(e^{\epsilon_n} - 1) - \kappa A_d(e^{\epsilon_n + \epsilon_d} - 1)$, where A_n and A_d are the integrated phase space areas over which the nonspecific and specific interactions of two rectangular monomers act (we find numerically that $A_n/A_d \approx 620$), and κ is a constant of proportionality with units of area. By rearranging this expression we see that lines of constant B_2 are straight lines of slope $1 - A_n/A_d$ on a plot of $(e^{\epsilon_n + \epsilon_d} - 1)$ versus $(e^{\epsilon_n} - 1)$. In Fig. S6 we have plotted, in such a manner, yield data from Fig. S5 (monomer concentration 1% by area) and from Fig. 3 (main text, monomer concentration 10.91% by area). The larger the square symbol size, the greater the yield of crystal found in dynamic simulations. Circles identify regions where low-quality- or no crystallization is seen. Dotted lines are of constant B_2 (they are curved because of the linear-log scale). Broadly, we find that crystallization is indeed seen within a fairly well-defined range of values of B_2 , but that model proteins with similar values of B_2 can behave differently: model proteins with a) strong specific attractions or b) moderately strong specific and nonspecific attractions may possess similar values of B_2 , but can crystallize with different fidelity (and via different dynamical pathways). Although the difference in yield as one moves along lines of constant B_2 is not always dramatic, this observation reveals that, at least for this model system, order parameters in addition to B_2 are required in order to understand its crystallization.

* Electronic address: swhitelam@lbl.gov

[1] B. Chen and J. I. Siepmann, *J. Phys. Chem. B* **105**, 11275 (2001).

[2] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press New York, 1987).

[3] A. George and W. W. Wilson, *Acta Crystallographica Section D: Biological Crystallography* **50**, 361 (1994).

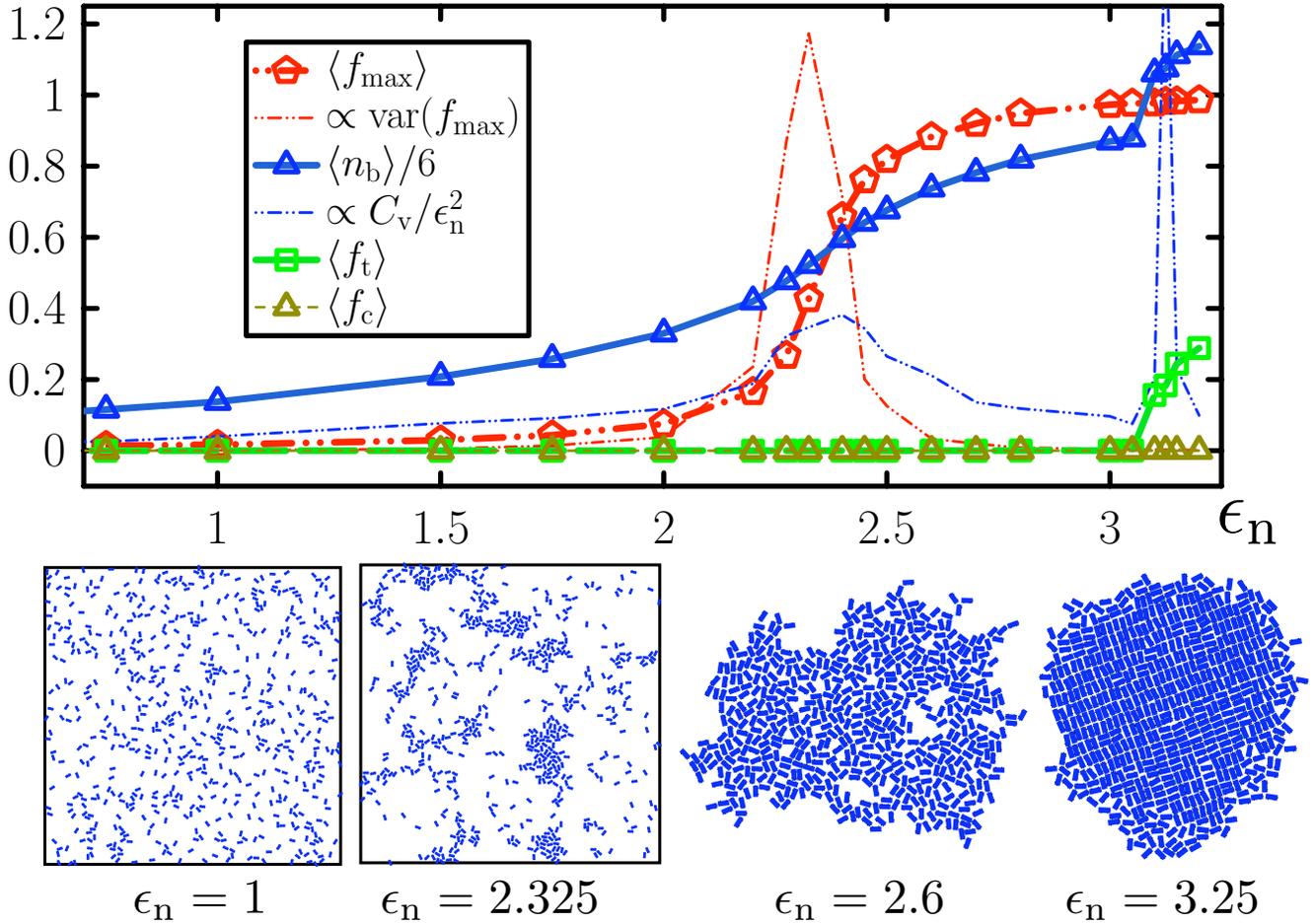


FIG. S1: Supplement to Fig. 1 (main text). Equilibrium behavior of monomers lacking specific interactions ($\epsilon_d = 0$) (600 particles, 10.91% coverage by area). Data points represent thermal averages, while lines are a guide to the eye. f_{\max} is the fraction of particles resident in the system's largest cluster (we show also the variance of f_{\max} scaled by an arbitrary factor); n_b denotes twice the number of pairwise bonds per monomer (C_v/ϵ_n^2 is proportional to the variance of this quantity); f_t is the tetratic order parameter defined in the Supporting Information; and f_c is the fraction of crystalline particles in the simulation box (particles making 3 directional bonds). The snapshots shown correspond to clustering; near-critical fluctuations; a dense liquid cluster; and a cluster exhibiting non-crystalline tetratic order, from left to right. We note that $\langle f_c \rangle = 0$ for all values of ϵ_n considered.

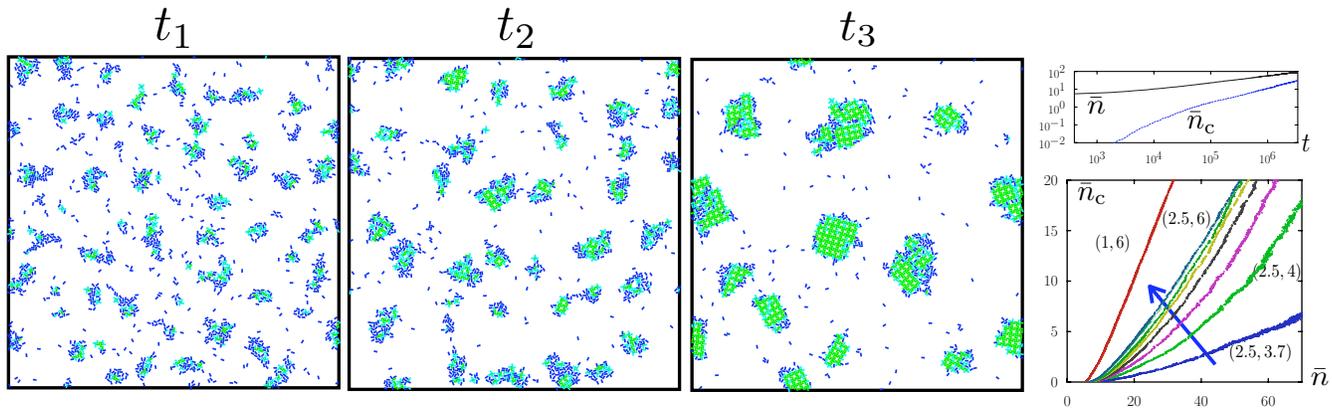


FIG. S2: Supplement to Fig. 2 (main text). Left: Time-ordered snapshots from a dynamic simulation (2000 monomers, 10.91% coverage by area) with $(\epsilon_n, \epsilon_d) = (2.5, 4)$. Crystallization corresponds to arrested liquid-vapor phase separation. Right: For the same model parameters we plot mean size \bar{n} and mean number of resident crystal particles \bar{n}_c of clusters of size 4 or larger as a function of number of MC steps t (top) and as a parametric function of t (bottom). In the bottom plot we show also parametric data from state $(\epsilon_n, \epsilon_d) = (1, 6)$ (where classical assembly is seen), and for states for which $\epsilon_n = 2.5$ and $\epsilon_d = 3.7, 4.25, 4.5, 4.75, 5, 5.5$ and 6. The arrow indicates the direction of increasing ϵ_d . While in this regime the liquidlike state is viable and metastable with respect to the crystal, its lifetime can be much reduced by increasing ϵ_d . Data averaged over 16 or 20 independent simulations.

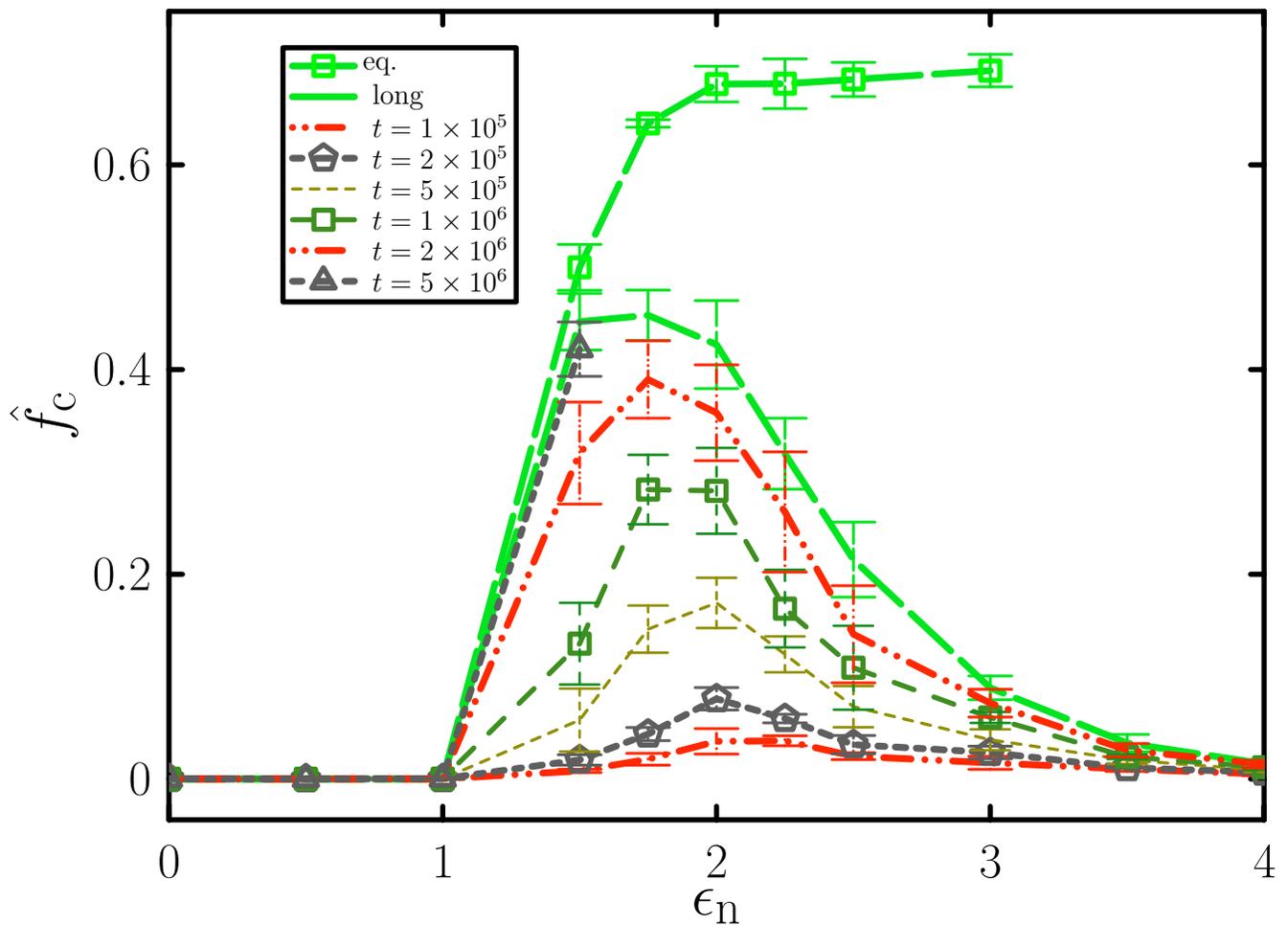


FIG. S3: Supplement to Fig. 3 (main text). Scaled crystal yields \hat{f}_c for fixed number of Monte Carlo steps for $\epsilon_d = 5$ and various fixed values of ϵ_n (600 particles, 10.91% coverage by area). Data points represent the mean of 5 independent simulations. We show also yields after fixed processor time (200 CPU hours per simulation, ‘long’) and equilibrium averages (‘eq.’). We draw similar qualitative conclusions of the basis of yields reported at specified processor times or specified numbers of MC steps.

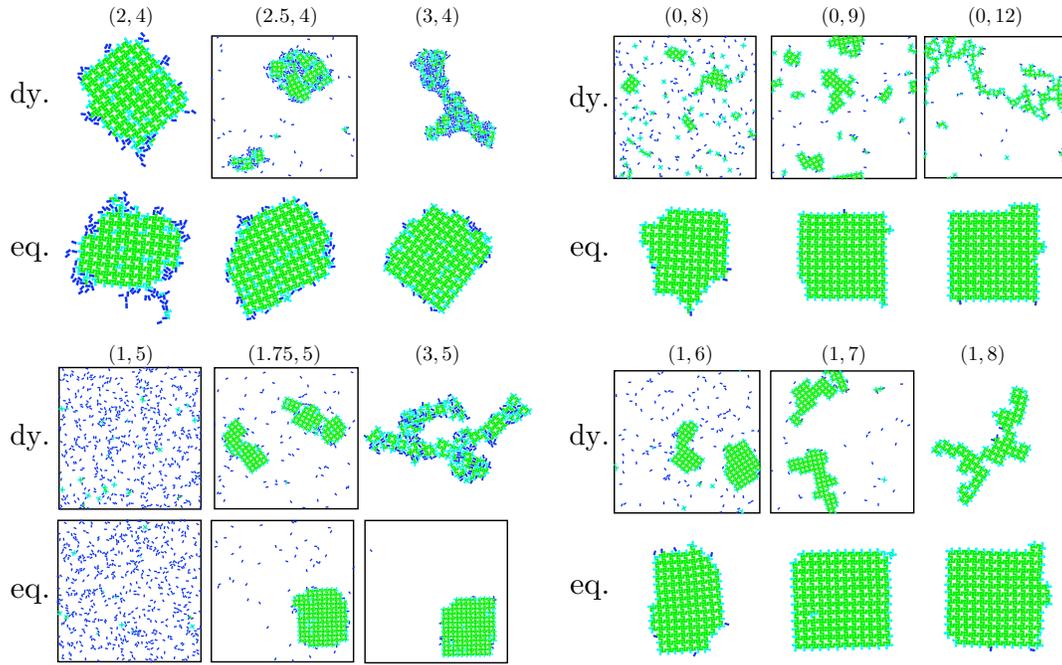


FIG. S4: Supplement to Fig. 3 (main text). Long-time configurations from dynamic simulations (dy.) and corresponding equilibrium configurations (eq.) (600 particles, 10.91% coverage by area). Snapshots are labeled (ϵ_n, ϵ_d) , and show either the whole simulation box or the largest cluster contained in the box.

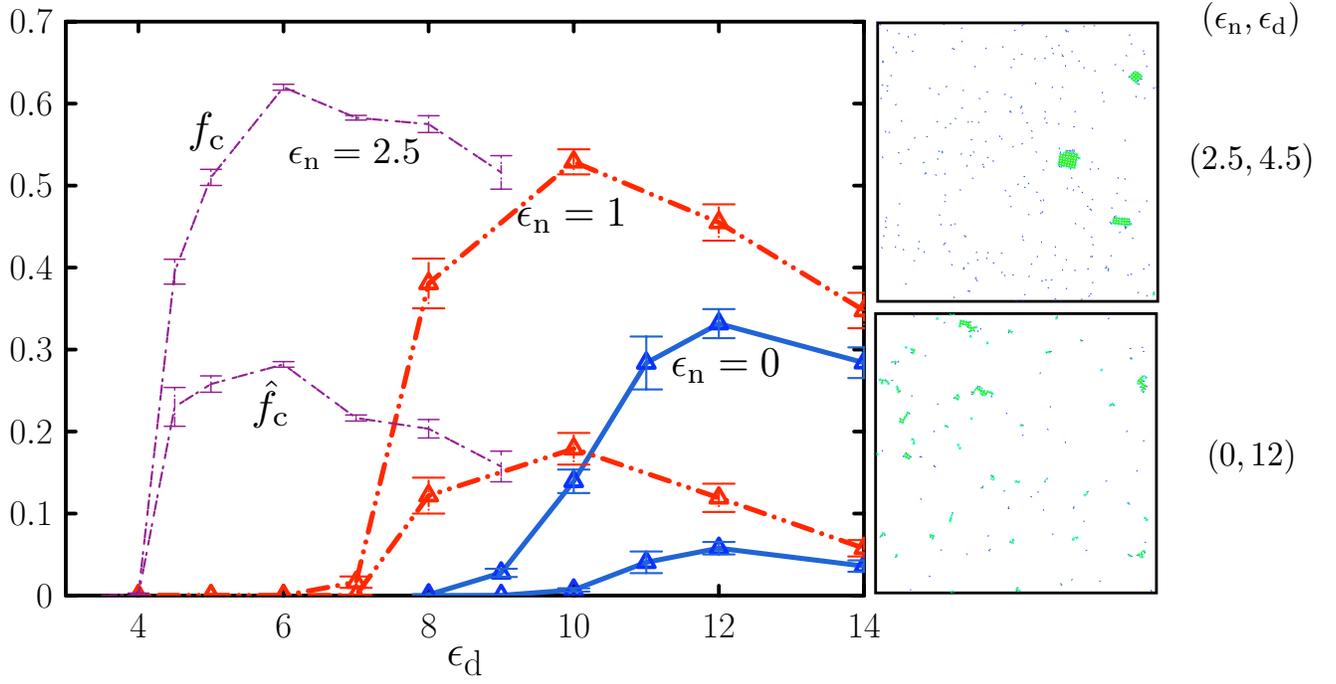


FIG. S5: Supplement to Fig. 3 (main text). Long-time crystal yields \hat{f}_c (lower three curves) and f_c (upper three curves) from simulations of 600 monomers present at concentrations of 1% by area. Similar behavior is seen to that shown in Figure 3 (main text), right panel (for 10.91% occupancy by area), with nonzero ϵ_n enhancing crystal assembly. At right: snapshots from two labeled states.

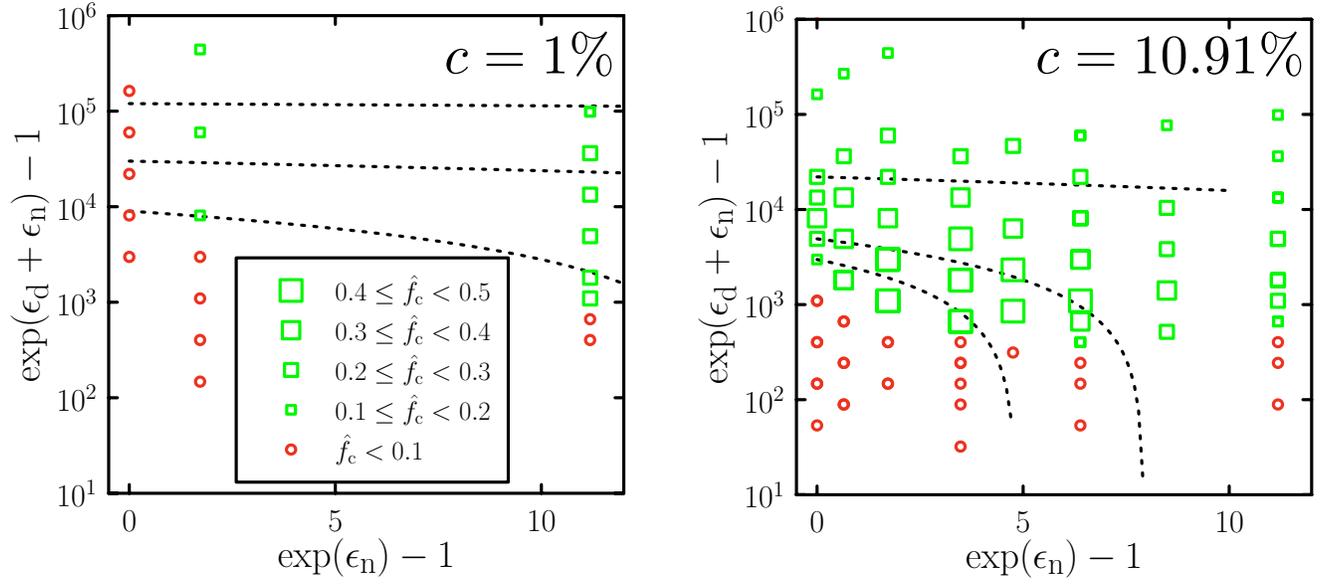


FIG. S6: Supplement to conclusions (main text). Monomers with similar values of B_2 can show different propensities for crystallization. We plot crystal yields from dynamic simulations for low concentration c of monomers (left panel, data from Fig. S5) and moderate concentration of monomers (right panel, data from Fig. 3, main text). Symbol size indicates crystal yield. Lines of constant B_2 are shown dotted. Along lines of constant B_2 , yield can be enhanced (and/or the dynamics of crystallization altered) by diminishing the strength of specific attractions and enhancing the strength of nonspecific interactions.