

# 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*

We use computer simulation to study crystal-forming model proteins equipped with interactions that are both orientationally specific and nonspecific. Distinct dynamical pathways of crystal formation can be selected by tuning the strengths of these interactions. When the nonspecific interaction is strong, liquidlike clustering can precede crystallization; when it is weak, growth can proceed via ordered nuclei. Crystal yields are in certain parameter regimes enhanced by the nonspecific interaction, even though it promotes association without local crystalline order. Our results suggest that equipping nanoscale components with weak nonspecific interactions (such as depletion attractions) can alter both their dynamical pathway of assembly and optimize the yield of the resulting material.

Controlling the crystallization of molecular and nanoscale systems remains a principal challenge of physics and chemistry. Controlling protein crystallization in particular is central to protein characterization, but despite advances in our understanding of protein phase behavior and association dynamics [1–15] we lack a set of rules for rational production of protein crystals *in vitro* [16]. Some proteins crystallize *in vivo*. S (‘surface’)-layer proteins form functional crystalline lattices on the outsides of many bacteria and archaea, and were among the first protein structures used to organize nanomaterials in a ‘bottom-up’ fashion [17, 18]. The sbpA S-layer protein from the bacterium *Lysinibacillus sphaericus* forms a square crystalline lattice of tetramers, and has been shown to crystallize in a ‘nonclassical’ fashion on supported lipid bilayers *in vitro* [19]: order emerges from dense amorphous clusters, rather than directly from crystalline nuclei. A similar dynamics is thought to operate during crystallization of the globular protein lysozyme [3, 4].

Here we introduce a molecular model designed to study crystallization in the presence and absence of amorphous intermediates. The model is inspired by the crystallization of the sbp S-layer, but is designed to be simple enough to allow us to draw conclusions about control of crystallization pathways more generally. The model comprises monomers equipped with two types of interaction. The first consists of a directionally nonspecific attraction, designed to mimic the tendency of proteins to associate in a manner that does not uniquely constrain the orientations of neighboring monomers. The second interaction comprises directionally- and chemically specific attractive patches whose placement is suggested by the S-layer’s electron density map [20] and its unusual crystal structure. Patches predispose monomers to the formation of a square crystalline lattice of tetramers. Here we attempt to answer the following question: How does the nonspecific interaction influence the dynamics of formation and yields of crystals whose symmetries are selected by the specific attraction?

In what follows we show that distinct dynamical path-

ways of crystal formation can be selected by tuning the strengths of nonspecific and specific interactions (this selection is suggested by the bulk free energy landscape of generic anisotropic particles [21], and by distinct dynamical pathways seen in simulation studies of virus capsid assembly [22] and polymer crystallization [23]). Nonclassical assembly via liquidlike intermediates is possible when the nonspecific interaction is strong; when it is weak, classical modes of assembly can be realized. In the former regime the lifetime of the liquidlike phase can be controlled by varying the strength of the specific interaction. We show also that optimal crystallization conditions are found when the nonspecific interaction is nonzero – a result striking in light of the fact that this interaction promotes none of the symmetries of the crystal – but not strong enough to induce the formation of large liquidlike intermediates. Other model proteins bearing both nondirectional and directional attractions have recently been studied, yielding valuable insight into phase behaviors and crystallization dynamics as temperature is varied [12, 15]. The present study is distinguished by its exploration of the dynamics and yields of assembly for nonspecific and specific interactions of varying absolute *and* relative strength. Such an exploration is required in order to assess monomers’ possible modes of assembly.

Model geometry is shown in Fig. 1. The model comprises a featureless two-dimensional substrate on which live, in continuous space, hard rectangular monomers of small edge length  $a$  and aspect ratio 2.2. Monomers possess both specific and nonspecific pairwise interactions. Specific interactions are mediated by three sticky patches placed on two sides of the rectangle, as shown, each a distance  $a/2$  from the nearest vertex. Patches are of type E (‘edge’), S (‘short-arm’) and L (‘long-arm’), and are selectively reactive: a ‘directional’ bond of energy  $-\epsilon_d k_B T$  is made when two L patches or one E- and one S patch are separated by a distance of less than  $a/5$ . Patch geometry predisposes monomers to the formation of a square lattice of tetramers, as sketched, in mimicry of the sbpA S-layer [20]. The tetrameric repeat unit of the latter measures about 18 nm on a side, and for correspondence we

imagine  $a \approx 4$  nm. In the simulations discussed below we defined particles making two directional bonds to be ‘partially crystalline’ (rendered light blue in snapshots), and particles making three directional bonds to be ‘crystalline’ (rendered green in snapshots). We denote by  $f_p$  and  $f_c$  the fractions of monomers in partially crystalline and crystalline states, respectively. The nonspecific interaction is a pairwise bond of energy  $-\epsilon_n k_B T$ , and is activated by the overlap of the two dotted rectangles shown; these rectangles are concentric with the monomers that give rise to them, and have side lengths  $2a/5$  in excess of the sides of those monomers. Interaction ranges assume solution conditions to be such that protein-protein interactions are attenuated on a length scale of about 1 nm.

We performed two types of *NVT* simulations within periodically-replicated, square boxes: ‘sampling’ and ‘dynamic’. Sampling simulations (designed to probe thermal equilibrium) employed 600 particles whose total area comprised 10.91% of the simulation box. Simulations were started from a configuration consisting of a square crystal (or a cluster of noncrystalline tetratic order) inserted into a vapor of monomers. We propagated these systems using local Monte Carlo moves supplemented by the nonlocal ‘teleportation’ algorithm described in the supplementary material [35]. Dynamic simulations were begun from configurations of randomly dispersed and oriented monomers, and were propagated using a ‘virtual-move’ Monte Carlo algorithm [24, 25]. Its purpose is to approximate a diffusive dynamics by using potential energy gradients to effect collective translations and rotations ignored by standard single-particle algorithms. Accounting for collective modes of motion is necessary in order to identify when a molecular system undergoing overdamped motion might assemble robustly or become kinetically trapped. We performed simulations of either 600 or 2000 monomers, and considered monomer occupancies by area ranging from 20% to 1% (focusing on the case of 10.91%). Further details of simulation protocol (and phase classifications) can be found in the supplementary material.

The phase diagram for the model in the  $(\epsilon_n, \epsilon_d)$  plane, derived from sampling simulations, is shown in Fig. 1. It identifies regimes of homogeneous fluid, phase-separated liquid and vapor, and crystalline order. The structure of the phase diagram is similar to that computed by mean field theory applied to prototypical anisotropic particles [21]: notably, liquid-vapor phase separation is in large part driven by the nonspecific interaction, and moderate values of this interaction enlarge the regime of crystal stability. For larger values of  $\epsilon_n$  we observe the emergence of a noncrystalline tetratic phase that owes its existence to monomers’ rectangular shape [26, 27] (see Fig. S1). Association driven by the nonspecific interaction stabilizes none of the order characteristic of the crystal: we find that  $\langle f_c \rangle = 0$  when  $\epsilon_d = 0$  (Fig. S1).

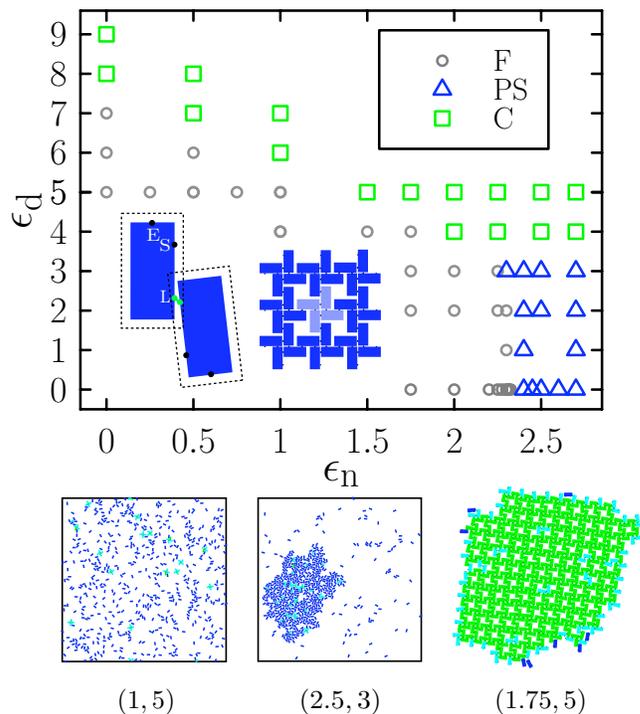


FIG. 1: Model geometry and phase diagram. Inset: Monomers consist of rectangles equipped with an attractive rectangular force-field (dotted) and decorated with three sticky patches labeled E, S and L. Only L-L and E-S pairings are reactive. Patch geometry predisposes monomers to the formation of a square lattice of tetramers (sketched), in mimicry of the sbpA S-layer. Main figure: Model phase diagram in the space of specific  $\epsilon_d$  and nonspecific  $\epsilon_n$  interaction strengths (600 particles, 10.91% coverage by area) shows regimes of homogeneous fluid (F), phase-separated liquid and vapor (PS), and crystal order (C). Snapshots below show examples of phases F, PS, and C, from left to right, taken from points  $(\epsilon_n, \epsilon_d)$ .

We next used dynamical simulation to determine how crystals form in different regions of parameter space. We found that crystallization can proceed by dynamical pathways both nonclassical – along which metastable liquidlike precursors form and only subsequently acquire crystalline order – and classical, along which the critical nucleus possesses the architecture of the stable solid. In Fig. 2 we show examples of both pathways. In general, the greater the value of  $\epsilon_n$  the greater the propensity for liquidlike clustering to precede crystallization (in a related vein, liquidlike clusters can precede the formation of model virus capsids if interaction patch specificities are sufficiently low [22]). If  $\epsilon_n$  is large enough to induce liquid-vapor phase separation then the resulting dynamics can comprise crystallization-arrested spinodal decomposition. This dynamics resembles that seen in experiment [19]. However, within this regime increasing  $\epsilon_d$  can shorten the lifetime of the metastable liquid by enhancing crystallization kinetics or (for large enough  $\epsilon_d$ ) inducing

assembly of gel-like intermediates (Fig. 2 and Fig. S2). This result is reminiscent of the kinetic stabilization of amorphous phases seen in computer simulations of liquid crystals [28], and suggests that in our model, as in that work, there exist regions of phase space within which Ostwald’s step rule does not hold. The latter states that the liquid phase, if stable with respect to the homogeneous fluid and metastable with respect to the crystal, should emerge prior to crystallization.

To determine the effect of the nonspecific interaction upon crystal quality we measured scaled yields  $\hat{f}_c \equiv f_c / (f_p + f_c)^2$  (an order parameter that rewards compact crystals with a large bulk-to-surface ratio) per particle after long dynamic simulations (Fig. S3) for fixed values of  $(\epsilon_n, \epsilon_d)$ . The left panel of Fig. 3 illustrates the effect of increasing  $\epsilon_n$  given  $\epsilon_d$ . For given  $\epsilon_d$  ( $\lesssim 9$ ), small values of  $\epsilon_n$  enhance assembly of the crystal, while large values induce dynamic arrest (cf. equilibrium behavior (inset); see also Fig. S4). The value of  $\epsilon_n$  at which arrest occurs is a function of  $\epsilon_d$ : in general, optimal assembly for given  $\epsilon_d$  occurs when  $\epsilon_n$  is too small to induce the formation of large liquidlike clusters, in accord with a suggestion made on the basis of a study of isotropic model proteins [14]. For certain choices of the specific interaction, however, such as  $\epsilon_d = 4$ , yields are maximized close to the liquid-vapor critical region. For larger  $\epsilon_d$  ( $> 9$ ), nonzero  $\epsilon_n$  provides little or no enhancement of yield. The right panel further reveals that the regime of best assembly occurs for small but nonzero values of  $\epsilon_n$ ; we observed similar behavior at monomer concentrations of 1% by area (Fig. S5). This enhancement of yield by the nonspecific interaction is striking in light of the fact that the latter promotes association without stabilizing the local order of the crystal. This result evokes one obtained from simulation studies of the self-assembly of closed virus capsids, namely, that capsid yield is optimized by interaction patch specificities that are neither too high nor too low [22, 29]. We speculate that in our model this enhancement has the following origin. Partial reversibility – the ability of components to transiently break bonds in order to correct the nascent structural defects of growing assemblies – is a necessary condition for robust self-assembly [30–32]. Particles bearing moderately strong nonspecific and specific interactions and particles equipped with very strong specific interactions may form solids of similar thermodynamic stability. However, it is likely that the former gives rise to a greater degree of ‘partial reversibility’ than does the latter: it is easier to break in sequence two moderately strong bonds than one very strong bond in order to correct nascent defects as structures grow. It is likely also that at very low monomer concentrations the increased collisional cross-section associated with the nonspecific interaction leads to an enhanced kinetics of assembly.

We have used computer simulation to study a model of crystal-forming monomers equipped with interactions

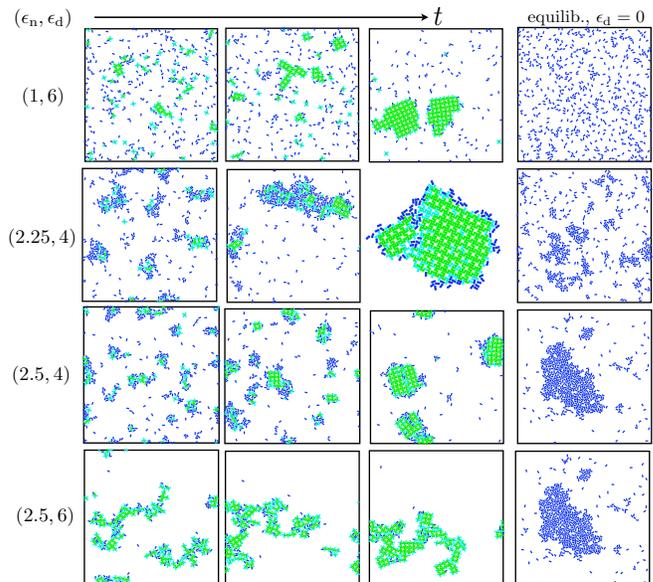


FIG. 2: Time-ordered snapshots from dynamic simulations (600 monomers, 10.91% coverage by area) for four different choices of  $(\epsilon_n, \epsilon_d)$ . Mechanisms of crystal assembly range from classical (top row), where the growing nucleus possesses the architecture of the stable solid, to nonclassical (rows 2 and 3), where the crystal emerges from the midst of dense liquidlike clusters. There also exist regimes (e.g. bottom row) in which the formation of gel-like networks prevents the emergence of the metastable liquid phase. At right: typical configurations in equilibrium in the absence of the specific interaction.

that are both nonspecific, in an orientational and chemical sense, and specific. Distinct dynamical pathways of crystal formation can be selected by tuning the strengths of these interactions. Fluctuations of density and structure sometimes cooperate (enhancing assembly), and sometimes conflict (impairing assembly). While both scenarios are suggested by simulations of isotropic model proteins [8, 14], here the presence of two types of interaction allows such fluctuations to be varied in strength (at fixed temperature and concentration) with a high degree of independence. We do not know the extent to which the qualitative findings of our solvent-free, two-dimensional simulations are relevant to real nanoscale components in three dimensions – where, for instance, reorganization of liquidlike intermediates might be considerably more rapid than in 2d – but direct extrapolation suggests that trading specific- for nonspecific interaction strength can alter assembly pathways and might be one way to optimize assembly. In protein solutions one could change the respective magnitudes of specific and nonspecific interactions by altering ionic strength and using inert nanoscale components to induce a depletion attraction [33]. Our results suggest that by trading specific- for nonspecific interaction strength, proteins with similar values of the second virial coefficient  $B_2$  can be made to crystallize

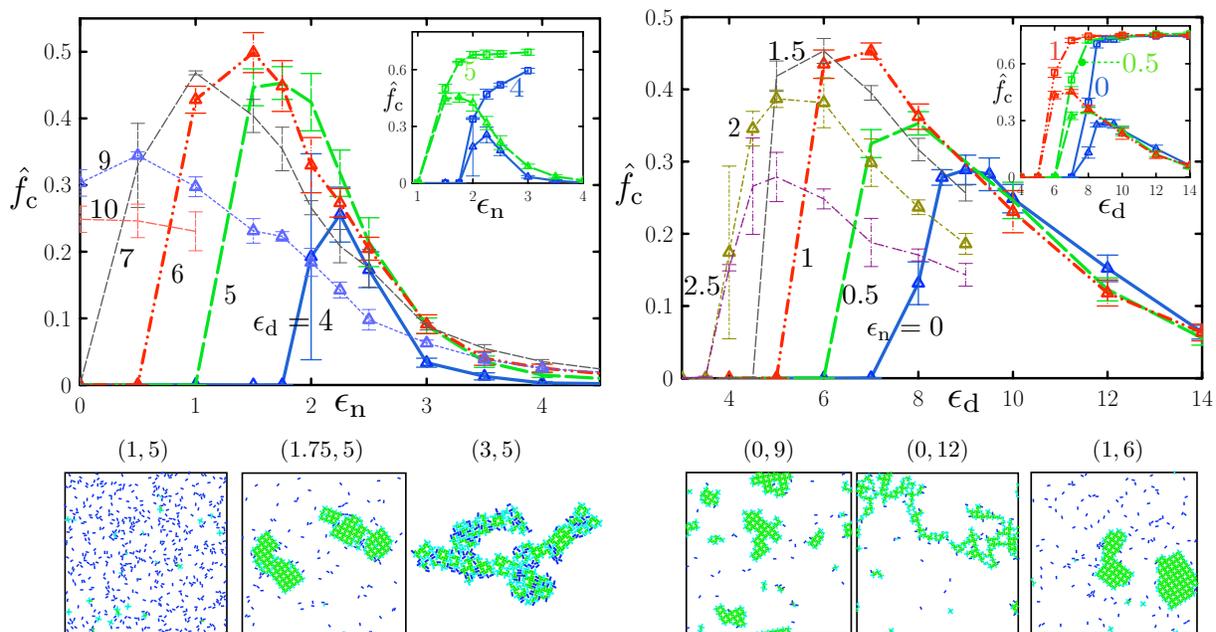


FIG. 3: Long-time scaled yields  $\hat{f}_c$  from dynamic simulations at specified fixed values of  $(\epsilon_n, \epsilon_d)$  (600 particles, 10.91% coverage by area; values of  $\epsilon_d$  and  $\epsilon_n$  label lines in left and right panels, respectively). Data points represent the mean of 5 independent simulations; lines are a guide to the eye. Insets compare selected sets of dynamic simulations with their equilibrium counterparts. The left panel shows the general enhancement of crystal yield conferred by nonzero  $\epsilon_n$ , for given  $\epsilon_d$  (up to  $\epsilon_d \approx 9$ ). The right panel shows that ‘best’ assembly is found in general for nonzero  $\epsilon_n$ , even though the nonspecific interaction stabilizes none of the symmetries of the crystal. Snapshots below show configurations from simulations at specified  $(\epsilon_n, \epsilon_d)$ .

in dynamically distinct ways and with different propensities (Fig. S6). This suggestion is consistent with the observation [34] that even proteins possessing values of  $B_2$  within the ‘crystallization slot’ are not guaranteed to crystallize.

We thank Sungwook Chung, Seong-Ho Shin, Jim DeYoreo, Carolyn Bertozzi and Caroline Ajo-Franklin for discussions. This work was performed at the Molecular Foundry, Lawrence Berkeley National Laboratory, and was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

\* Electronic address: [swhitelam@lbl.gov](mailto:swhitelam@lbl.gov)

- [1] M. Muschol and F. Rosenberger, *J. Chem. Phys.* **107**, 1953 (1997).
- [2] R. P. Sear, *J. Chem. Phys.* **111**, 4800 (1999).
- [3] P. G. Vekilov, *J. Crystal Growth* **275**, 65 (2005).
- [4] O. Galkin and P. G. Vekilov, *Proc. Nat. Acad. Sci.* **97**, 6277 (2000).
- [5] G. Foffi, G. D. McCullagh, A. Lawlor, E. Zaccarelli, K. A. Dawson, F. Sciortino, P. Tartaglia, D. Pini, and G. Stell, *Phys. Rev. E* **65**, 31407 (2002).
- [6] L. F. Filobelo, O. Galkin, and P. G. Vekilov, *J. Chem. Phys.* **123**, 014904 (2005).
- [7] W. Pan, A. B. Kolomeisky, and P. G. Vekilov, *J. Chem. Phys.* **122**, 174905 (2005).
- [8] P. R. ten Wolde and D. Frenkel, *Science* **277**, 1975 (1997).
- [9] R. P. Sear, *J. Phys. Cond. Mat.* **19**, 033101 (2007).
- [10] J. P. K. Doye, A. A. Louis, I. C. Lin, L. R. Allen, E. G. Noya, A. W. Wilber, H. C. Kok, and R. Lyus, *Phys. Chem. Chem. Phys.* **9**, 2197 (2007).
- [11] H. Liu, S. K. Kumar, and F. Sciortino, *J. Chem. Phys.* **127**, 084902 (2007).
- [12] C. Gögelein, G. Nägele, R. Tuinier, T. Gibaud, A. Stradner, and P. Schurtenberger, *J. Chem. Phys.* **129**, 085102 (2008).
- [13] S. Auer, C. M. Dobson, M. Vendruscolo, and A. Maritan, *Phys. Rev. Lett.* **101**, 258101 (2008).
- [14] B. Chen, R. B. Nellas, and S. J. Keasler, *J. Phys. Chem. B* **112**, 4725 (2008).
- [15] H. Liu, S. K. Kumar, and J. F. Douglas, *Phys. Rev. Lett.* **103**, 18101 (2009).
- [16] L. Slabinski, L. Jaroszewski, A. P. C. Rodrigues, L. Rychlewski, I. A. Wilson, S. A. Lesley, and A. Godzik, *Protein Science* **16**, 2472 (2007).
- [17] P. Messner and U. Sleytr, *Adv. Microbial Physiology* **33**, 213 (1992).
- [18] U. B. Sleytr, *FEMS Microbiology Reviews* **20**, 5 (1997).
- [19] S. Chung, S. H. Shin, C. Bertozzi, and J. De Yoreo, Submitted (2010).
- [20] J. E. Norville, D. F. Kelly, T. F. Knight, A. M. Belcher, and T. Walz, *J. Struct. Biol.* **160**, 313 (2007).
- [21] S. Whitelam, *J. Chem. Phys.* **132**, 194901 (2010).
- [22] A. W. Wilber, J. P. K. Doye, A. A. Louis, E. G. Noya, M. A. Miller, and P. Wong, *J. Chem. Phys.* **127**, 085106 (2007).

- (2007).
- [23] W. Hu and D. Frenkel, Interphases and Mesophases in Polymer Crystallization III pp. 1–35 (2005).
- [24] S. Whitelam and P. L. Geissler, *J. Chem. Phys.* **127**, 154101 (2007).
- [25] S. Whitelam, E. H. Feng, M. F. Hagan, and P. L. Geissler, *Soft Matter* **5**, 1251 (2009).
- [26] J. Geng and J. V. Selinger, *Phys. Rev. E* **80**, 11707 (2009).
- [27] A. Donev, J. Burton, F. H. Stillinger, and S. Torquato, *Phys. Rev. B* **73**, 54109 (2006).
- [28] O. Henrich, K. Stratford, D. Marenduzzo, and M. E. Cates, *Proc. Natl. Acad. Sci. USA* **107**, 13212 (2010).
- [29] M. F. Hagan and D. Chandler, *Biophys. J.* **91**, 42 (2006).
- [30] G. M. Whitesides and M. Boncheva, *Proc. Nat. Acad. Sci.* **99**, 4769 (2002).
- [31] R. L. Jack, M. F. Hagan, and D. Chandler, *Phys. Rev. E* **76**, 21119 (2007).
- [32] D. Rapaport, *Phys. Rev. Lett.* **101**, 186101 (2008).
- [33] D. Marenduzzo, K. Finan, and P. R. Cook, *J. Cell Biol.* **175**, 681 (2006).
- [34] A. George, Y. Chiang, B. Guo, A. Arabshahi, Z. Cai, and W. W. Wilson, *Methods in Enzymology* **276**, 100 (1997).
- [35] Linked from this article's arXiv page.