

The role of non-specific interactions in a patchy model of protein crystallization

Iskra Staneva¹ and Daan Frenkel^{1, a)}

*Department of Chemistry, University of Cambridge
Lensfield Road, Cambridge CB2 1EW, United Kingdom*

(Dated: 22 November 2015)

We use a coarse-grained model for generic proteins to investigate the formation of structures with $P2_12_12_1$ symmetry, the most prevalent space group of protein crystals. To account for the string directionality of protein-protein interactions that has been suggested by previous studies, we represent proteins as spherical particles that are covered by a large number of small, attractive ‘patches’ that are randomly distributed on the protein surface. Attractive interactions between two proteins can then involve several pairs of patches interacting simultaneously. Our results suggest that the unit cell with the lowest energy is not necessarily the one that grows fastest. Rather, growth is favoured if 1) new particles can attach with enough bonds to the growth front and 2) particles that attach in crystallographically inequivalent positions bind to the surface with similar strength. We subsequently study the impact of interactions that are not part of crystalline contacts, and find that when these non-specific interactions are few and weaker than the crystal contacts, both nucleation and growth are successful. If the proportion of non-specific interactions is increased, crystal growth is still possible in a small range of model temperature.

I. INTRODUCTION

Making high-quality protein crystals is crucial for successful protein structure determination by X-ray diffraction. Moreover, the control of protein crystal morphology is increasingly important for the formulation of pharmaceutical compounds¹. Yet very little can be predicted about the crystallization process, and in practice, growing high-quality protein crystals proceeds through trial-and-error involving a scan of a wide range of protein solution conditions. Questions concerning, for instance, the role of co-solutes, the properties of possible intermediate phases, and the kinetics of nucleation and growth are still not fully answered.

If these questions are to be investigated through computational means, the main difficulty is due to the large sizes of these molecules (even a fairly small protein like lysozyme contains ~ 1000 atoms). This makes long enough large-scale simulations using an atomistic representation unfeasible. While there are calculations of protein solution properties performed with 500-1000 atomically detailed rigid molecules^{2,3}, theoretical studies of the full protein phase behavior typically require hundreds, if not thousands of different simulations. This can at present, for systems containing hundreds of proteins, only be carried out using coarse-grained models, where globular proteins are portrayed as colloidal spheres, and their interactions are mediated by short-range isotropic potentials⁴⁻⁷ or attractive spots⁸⁻¹³. Such descriptions have, for example, successfully managed to generate phase diagrams consistent with experimental measurements^{5,6,8,11}, explained re-entrant condensation of proteins whose interactions are ion-activated¹³, and proposed how nucle-

ation barriers might depend on the distribution of attractive patches on the molecular surface^{12,14}.

Here we use a patchy-sphere model to address fundamental questions related to crystal formation, and to study the relative importance of specific and non-specific interactions for the crystal growth process. Previous theoretical studies typically designed patch positions, size and amounts such that only one or two¹⁵ single crystal structures had the possibility to assemble. The model that we present here is based on spheres decorated completely randomly with a large number of small attractive patches, portraying the fact that proteins have large surface areas allowing for several non-specific interactions. Intermolecular contacts that consist of more than one patch-pair are also taking into account the directionality of protein-protein interactions¹⁰. This random patch positioning enables us to explore the possibility of finding a unit cell without *a priori* stipulating a structure. A similar model was recently shown to be able to fit liquid-liquid phase separation curves and osmotic compressibilities to experimental data¹⁶.

We focus on structures with $P2_12_12_1$ symmetry, since this is the most prevalent space group of protein crystals, and find that proteins represented in this fashion can crystallize only if the crystal contacts are strong enough compared to the non-specific interactions. The larger the number of non-specific interactions, the narrower is the temperature interval where crystals can grow. Similarly to supersaturation conditions, the region in model parameter space for successful crystal nucleation and the one for growth are not entirely identical. Finally we observe that a unit cell with the same energy as the $P2_12_12_1$ structure but of another symmetry does not correspond to a crystal capable of growing. We suggest that this could be due to a large spread in the interaction energies that different crystal layers offer to attaching particles.

^{a)} Electronic mail: df246@cam.ac.uk

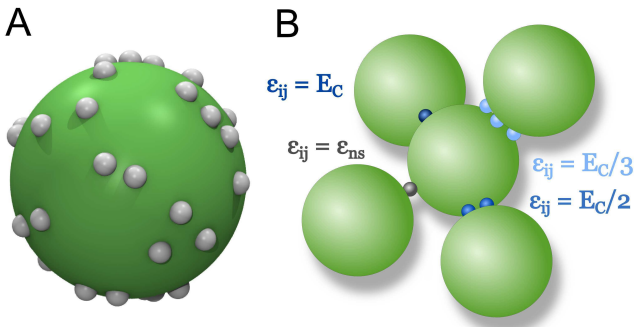


FIG. 1. **Patchy protein model.** (A) An illustration of the coarse-grained protein model with $N_p = 40$ patches on the surface, each represented as a sphere with a diameter that corresponds to the range of interaction (0.1σ , where σ is the protein diameter). (B) An example of the assignment of ε_{ij} for specific interactions when the crystal contacts are formed by one (dark blue), two (lighter blue) and three (lightest blue) patch-pairs. For an interacting patch-pair that is not part of a crystal contact (represented with gray in the figure), $\varepsilon_{ij} = \varepsilon_{ns} = E_C/\xi$.

II. METHODS

A. Protein model

Proteins are modeled as hard spheres of diameter σ , interacting with each other through patches on their surface (see figure 1A). The N_p patches of a protein are randomly distributed, and a patch i on a molecule interacts with patch j on another molecule with a square well potential:

$$E_{ij}(r) = \begin{cases} -\varepsilon_{ij} & \text{if } r < 0.1\sigma \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where r is the distance between the patches. All particles in a simulation have identical patch decorations.

During the initial scan for potential crystal structures ε_{ij} were set equal (non-specific) for all combinations of patch-pairs, *i.e.* $\varepsilon_{ij} = \varepsilon_{ns} = 1kT$ for all i, j . For this choice of parameters, we measured a distribution of molecular pairwise interaction minima (see figure 2). We note that our ‘micro-patch’ model qualitatively reproduces the distribution of interaction energies between pairs of atomistically modeled proteins, as studied by Quang *et al.*¹⁷. In particular, it shows the same long-tail shape. This similarity is obtained when N_p (or the patch width) is large enough. In order to facilitate an efficient execution of the simulations, and to be able to vary patch specificity later on, in what follows N_p is chosen to be 40.

Once a structure was chosen for the study of the impact of non-specific interactions on crystal nucleation and growth, two parameters, ξ and η , were introduced to regulate the competition between correct crystalline bonds and improper associations. η , a parameter that can as-

sume values between 0 and 1, sets the amount of non-specific interactions present by allocating ηN_p possible binding partners per patch. Aside from patch-pairs that are identified as part of a crystal contact, the remaining patches are paired randomly until the correct number of interactions is reached.

ξ establishes the relative strength between specific and non-specific interactions and is defined as follows: if patches i, j do not participate in a crystalline contact, $\varepsilon_{ij} = \varepsilon_{ns}$. For simplicity, and since it has been shown that high asymmetry in crystal contact energies impedes crystallization thermodynamics and kinetics¹², we keep all crystal contacts equally strong, independently of how many patch-pairs they comprise. In each crystal contact, the contact energy E_C is divided equally among the constituting patch-pairs (see figure 1B). ξ is then defined as the ratio between the energies of a crystal contact and the non-specific patch-pair interaction, *i.e.* $\xi = E_C/\varepsilon_{ns}$.

In general, the choice of equal strength of crystal contacts may seem to be a strong constraint. However, this is where the special feature of the $P2_12_12_1$ crystal lattice comes in: the position and orientation of particles in the unit cell of such lattices is characterized by 7 independent parameters¹⁸. Hence, the ‘space’ of possible $P2_12_12_1$ unit cells is very large. We argue that the unit cells that form are precisely the ones where all crystal contacts are of comparable strength.

B. Simulation techniques

The search for symmetric low-energy configurations was done with the Monte Carlo variable box shape method described by Filion *et al.*¹⁹. A unit cell with periodic boundary conditions containing 4 particles (since this number is sufficient to represent the most frequent space groups of protein crystals¹⁸) was simulated, where the particles’ positions and orientations were varied, as well as the shape and size of the cell. The symmetry of the configurations was checked using the software of Stokes and Hatch²⁰. The crystal growth was studied with Monte Carlo (MC) NVT simulations where a solution of 100 monomers was brought into contact with a crystalline substrate with a thickness of one unit cell. The particles of the preformed layer were free to rotate, but were bound to their ideal crystal sites with harmonic potentials. A particle was considered as crystalline if it had made correct bonds to at least two other crystalline particles. For each examined crystal structure, growth was simulated in the x , y and z directions.

Results of the NVT computations were validated against more expensive simulations in the grand-canonical ensemble (constant μVT), where the systems were prepared as described previously, and μ was chosen to give a density of particles that allowed for growth of the crystal layer.

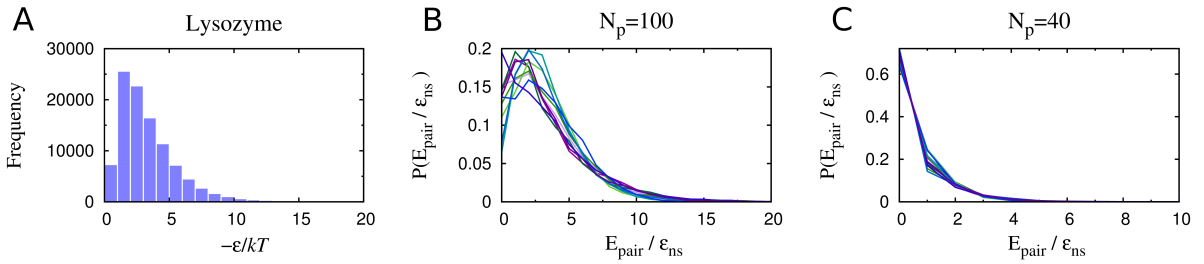


FIG. 2. **Distribution of pairwise energies.** Sampled energies between two particles for different relative orientations. Comparison is made between atomistically represented lysozyme (A) and our model (B and C), for particles with $N_p = 100$ (B) and $N_p = 40$ (C) patches. The different curves are for different patch decorations. The data values in (A) are extracted from figure 1 in Quang *et al.*, *J. Chem. Theory. Comput.*, vol 10, pp. 835-845 (2014).

III. RESULTS

A. Searching for symmetric low-energy structures

We analyzed the lowest-energy structures rendered with the variable box shape method, for several different patch decorations. Our results showed that they did not possess any symmetry other than the one of space group P1. The search was therefore narrowed down by constraining the generated configurations to only assume $P2_12_12_1$ symmetry. This choice was made because this is the most prevalent space group for proteins, found in $\sim 30\%$ of the crystals (see figure S1²¹).

The introduced constraints implied keeping the simulation box orthorhombic and relating the relative positions and orientations of the particles to the relevant symmetry operators. This was done by modifying the variable box method so that only the lengths of the vectors spanning the unit cell were varied. Furthermore, once the first particle was placed in the unit cell with a random position and orientation, the $P2_12_12_1$ symmetry operators were applied on the coordinates of its center of mass and those of its patches to calculate the positions and orientations of the remaining three particles. Only unit cells that, when repeated, would form fully bonded three-dimensional networks were considered.

This second set of generated conformations spanned a wide range of energies and densities (see figure S2). Such continuous distribution of unit cell energies does not single out one particular candidate structure for subsequent growth calculations. Importantly, simply choosing the configurations with minimum energy proved not to be sufficient for determining the crystal that will grow successfully. This was mainly because of the large variation in crystal bond strengths, that would lead to situations where one particle layer readily binds, while the next one does not.

The same procedure was repeated using particles with a lower number of patches, $N_p = 15$. As before, there was not one single structure with distinguishable low energy. Configurations with the lowest energies could not show successful growth for all layers or growth directions. To

test whether growth was inhibited by non-specific interactions, the particles were stripped of all their patches apart from the ones participating in crystal contacts. One patch-pair per contact was kept, and the range of attraction of the patches was doubled. Even with all non-specific interactions removed, the structure failed to grow. The reason was that the initiation of every second layer required the first particle to attach through one single bond. Growth could not be achieved by tuning the strength of the crystal contacts. Choosing a lower $|\epsilon_{\text{ns}}|$ prevented the particles from attaching at all, while increasing $|\epsilon_{\text{ns}}|$ did not resolve the problem of distinguishing the formation of correct patch pairing from the incorrect ones. Our findings thus suggest that crystal growth is facilitated if the attaching particle can form bonds with several members of the underlying layer, and furthermore, if not all patches have equivalent interactions.

B. A successful structure

Inspired by the observations reported above, we turned our focus to finding a configuration where each new particle would attach with a maximum amount of crystal bonds, no matter whether the crystal was grown in the x , y or z direction. One such structure was found, where in each direction each particle was making at least 4 bonds upon attachment to the crystal layer. The particles in this structure had 10 contacts each, consistent with the 8 – 10 molecular pairwise interfaces observed in protein crystals²². The bonds were formed out of single (four of the bonds), double (two of the bonds) and triple patch-pairs.

C. Growth dependence on relative interaction strength and specificity

Having identified an adequate crystal structure, we used it to explore how non-specific interactions influence crystal growth. Growth simulations were performed

varying the specificity and the crystal contact strengths, represented by η and ξ , respectively. Figure 3 shows the fraction of particles that have attached correctly to the crystal layer at the end of simulations of length $\sim 3 \cdot 10^7$ MC cycles. As can be seen, below $\xi \approx 2.1$ nothing attaches, and in some cases the layer is itself not stable (the crystallinity is lost due to rotation of the layer particles). Note however that at $\xi = 2.7$ crystal growth is possible, even though it means, for this particular structure, that ε_{ij} of some crystal-related patch-pairs (those that form a contact of valency three) is lower than the non-specific interactions.

Best growth is found in a region with high ξ and low η (strong crystal contacts and few interfering interactions). However, as seen in figure 4, if the non-specific interactions, ε_{ns} , are lowered and the crystal contacts are strong enough, crystal growth is possible also for cases when $\eta = 1.0$, *i.e.* when the amount of interfering, non-crystalline interactions is maximal. This is consistent with the observation of George and Wilson²³ that protein crystallization often occurs in conditions characterized by a slightly negative osmotic second virial coefficient, *i.e.* modest attraction between the proteins. In this case modest attraction is obtained through either a few non-specific patch-pairs, or through a larger amount but with weaker interactions. Furthermore, for a fixed E_C , the range of non-specific interaction strengths where crystal growth is observed is larger for lower η compared to high ones (figure 5).

D. Possibility to grow other structures

The patch-patch interaction matrices for a set of the η, ξ combinations that rendered successful crystal growth were used to check if unit cells of lower energies can be built. Attempts were made to build new structures of $P2_12_12_1$ symmetry, as well as structures with $P2_1$ symmetry (the second most common space group of protein crystals). When looking for other $P2_12_12_1$ structures and using ε_{ij} from the high- ξ /low- η region, the lowest-energy configurations found were very similar to the original unit cell (crystal contacts differing by only a few patch-pairs). Applying instead the interaction parameters used at $\varepsilon_{ns} = 2 kT, \eta = 1.0, \xi = 3.0$, new configurations featuring the same energy per particle as the original structure were generated from both types of space groups, but they did not grow.

A possible explanation for this inability could be that the tested structures displayed a larger variation in interaction energies between the different crystalline surfaces and particles commencing a new layer. As presented in figure 6, even if the average attachment energy for one of the new configurations (denoted “ $P2_12_12_1$ ”) is lower than the original structure, the standard deviation of this computed mean is larger.

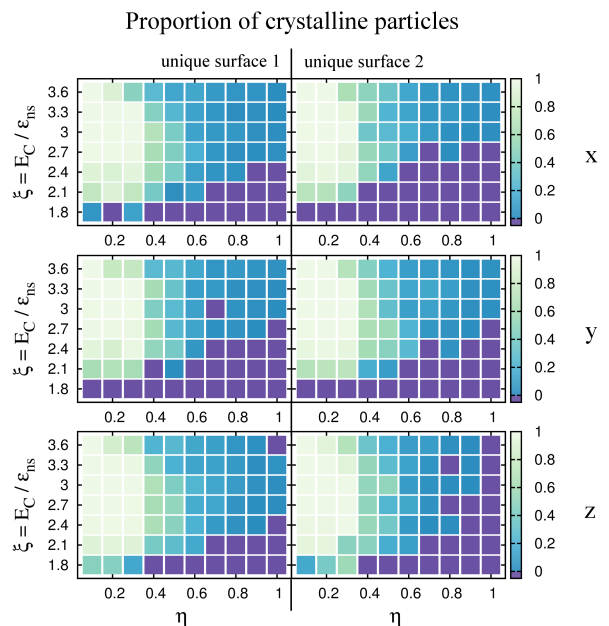


FIG. 3. **Crystal growth as a function of crystal contact strength and amount of interfering interactions.** The proportion of crystallized particles at the end of growth simulations for $\varepsilon_{ns} = 3.0 kT$. The six different images correspond to three different growth directions (x , y and z), each initialized with two unique crystal layer surfaces, obtained by periodic shifts within the unit cell. Violet indicates that there are particles in the initial crystal layer that have lost their correct orientations. Blue to green specifies a stable and growing crystal layer. The more light green the color, the larger amount of particles have attached correctly.

E. Nucleation dependence on relative interaction strength and specificity

The possibility to nucleate a crystal from solution, without any crystalline layer present, was examined for the same range of η and ξ as in the growth studies. Figure 7 shows that the monomers can only crystallize for high ξ , when the crystal contacts are much stronger than the non-specific interactions. For $\eta \lesssim 0.4$ and $\xi \gtrsim 3.3$ the particles crystallize into a few, compact nuclei, while for larger η several crystallites form. However, as inferred from figure 8, which shows the ratio between the sizes of the largest assembly and the largest crystalline cluster, these smaller domains are generally part of larger aggregates interspersed with non-crystalline particles.

When ξ instead is low and $\eta \gtrsim 0.4$ the particles form disordered dynamic aggregates, with few or no crystalline particles (see figure 9). Examples of final configurations for different values of η and ξ are demonstrated in figure 10. Comparing figures 3 and 7, it is notable that in the region of $\eta \leq 0.3$ and low ξ where growth of a pre-formed layer is possible, nucleation does not occur. This is analogous to how a higher supersaturation is needed for nucleation compared to crystal growth²⁴. As figure 9

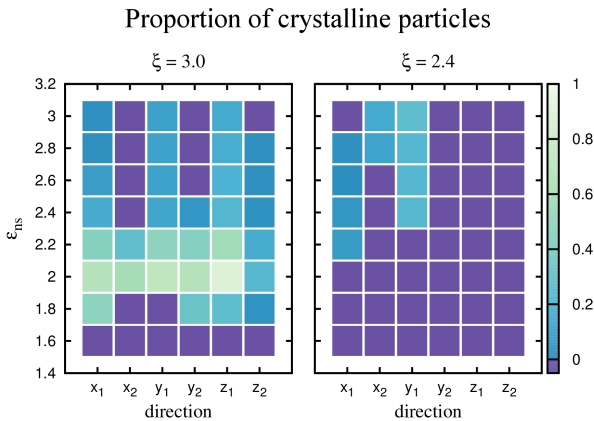


FIG. 4. **Possibility for crystal growth with a maximum amount of non-specific interactions.** The proportion of particles that have crystallized at the end of growth simulations when $\eta = 1.0$, as a function of ε_{ns} for $\xi = 3.0$ (left) and $\xi = 2.4$ (right). The six columns represent three different growth directions, each with two unique surfaces obtained through periodic shifts within the unit cell. Violet indicates that there are particles in the initial crystal layer that have lost their correct orientations. Blue to green specifies a stable and growing crystal layer. The more light green the color, the larger amount of particles have attached correctly.

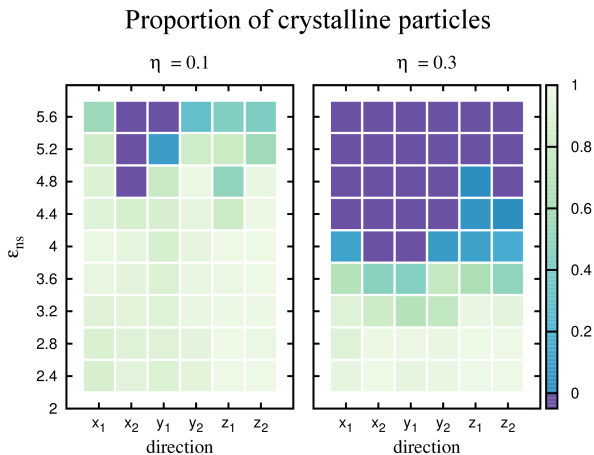


FIG. 5. **Ranges for crystal growth at a specific crystal contact strength.** The proportion of particles that have crystallized at the end of growth simulations when $E_C = 7.2 kT$, as a function of ε_{ns} for $\eta = 0.1$ (left) and $\eta = 0.3$ (right). The six columns represent three different growth directions, each with two unique surfaces obtained through periodic shifts within the unit cell. Violet indicates that there are particles in the initial crystal layer that have lost their correct orientations. Blue to green specifies a stable and growing crystal layer. The more light green the color, the larger amount of particles have attached correctly.

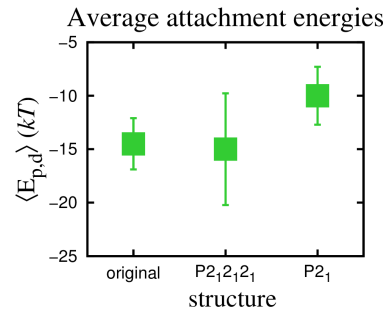


FIG. 6. **Comparison of attachment energies.** Energies for correctly attaching the first particle of a new layer, averaged over layer formation in three directions and over all particles in a unit cell. Values are calculated for the successfully growing structure (“original”) at $\varepsilon_{ns} = 2 kT$, $\eta = 1.0$, $\xi = 3.0$, and for two other structures that use the same patch-patch interaction matrix and have the same energy per particle: another configuration with $P2_12_12_1$ symmetry (“ $P2_12_12_1$ ”), and one from the $P2_1$ space group (“ $P2_1$ ”).

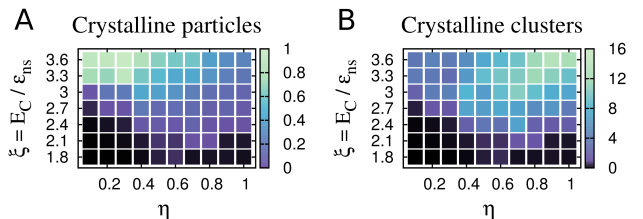


FIG. 7. **Crystal nucleation as a function of crystal contact strength and number of interfering interactions.** The proportion of crystalline particles (A) and number of crystalline clusters (B) at the end of nucleation simulations.

shows, instead of crystalline nuclei, only small, mainly disordered clusters containing $\lesssim 8$ particles form. These aggregates are transient and do therefore not provide a nucleus for further growth. The same observation is made for the state with $\eta = 1.0$ and $\xi = 3.0$. While such particles, as described before, correctly attach to a crystal layer in the narrow range $\varepsilon_{ns} = 2 - 2.2 kT$, no nucleation was observed for these conditions over the duration of the simulations. As figures 9B and S3 illustrate, at $\varepsilon_{ns} = 2 kT$ the particles do not cluster much, while at $\varepsilon_{ns} = 2.2 kT$ larger aggregates do assemble, but they are predominantly disordered.

IV. DISCUSSION

There has been some discussion about the suitability of colloidal models for studies of protein solution properties²⁵. In this study we consider that keeping this simplification is admissible, as it is assumed that the proteins are fully folded under crystallization conditions, and the questions addressed are related to crystal growth.

Our finding that crystallization is swiftest when the

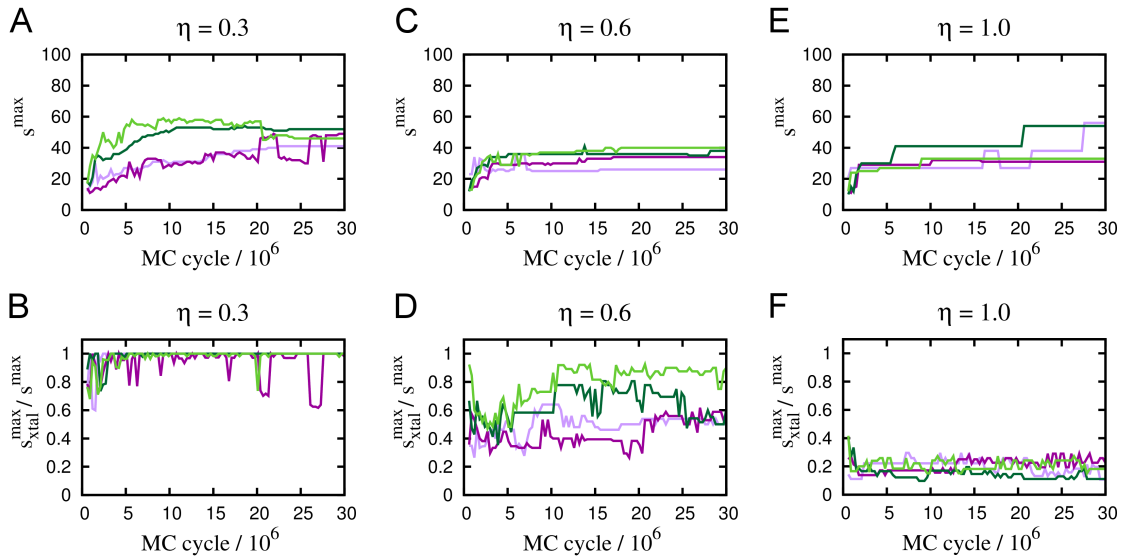


FIG. 8. **Cluster formation for different amounts of interfering interactions.** s^{max} , the size of the largest cluster of particles, shown for $\eta = 0.3$ (A), $\eta = 0.6$ (C) and $\eta = 1.0$ (E). Different colors represent different simulations. This is compared to the size of the largest crystalline cluster, s_{xtal}^{max} , as the fraction s_{xtal}^{max}/s^{max} for the same set of values of η in (B), (D) and (F), respectively. In all cases $\varepsilon_{ns} = 3kT$ and $\xi = 3.6$.

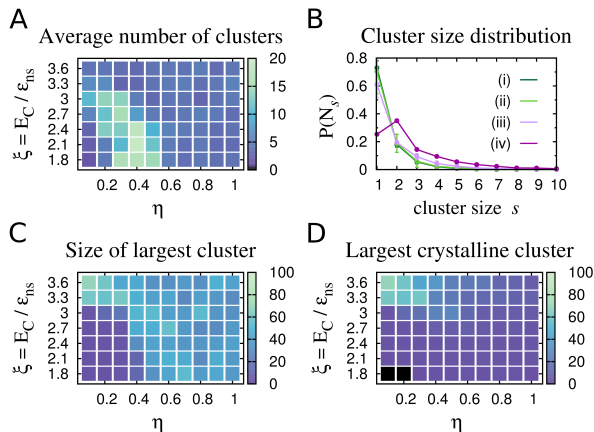


FIG. 9. **Cluster statistics of nucleation simulations.** (A) Average number of clusters of any type of structure, as a function of η and ξ at $\varepsilon_{ns} = 3.0kT$. (B) Probability for a particle to be part of a cluster of size s , calculated for regions in parameter space where growth is observed but not nucleation. The different curves correspond to the following parameters: (i) $\varepsilon_{ns} = 3.0kT, \eta = 0.1, \xi = 3.0$; (ii) $\varepsilon_{ns} = 3.0kT, \eta = 0.2, \xi = 2.7$; (iii) $\varepsilon_{ns} = 3.0kT, \eta = 0.3, \xi = 2.7$; (iv) $\varepsilon_{ns} = 2.0kT, \eta = 1.0, \xi = 3.0$. (C) The size of the largest aggregate as a function of η and ξ at $\varepsilon_{ns} = 3.0kT$. (D) The size of the largest crystalline cluster as a function of η and ξ at $\varepsilon_{ns} = 3.0kT$.

non-specific interactions are distinctively weaker compared to the crystal contact energies is similar to what was concluded from studies of 2D crystallization^{26,27}. Here we did not only study the effect of varying the

strength of the non-specific interactions, but also the proportion of them, observing that crystal growth is achievable also in the presence of a substantial amount of possibilities to form non-crystalline bonds. Since this case is presumably most similar to real proteins, the effects of the non-specific interactions, *i.e.* narrowing the temperature interval where a crystal can grow, concurrent with hampering nucleation due to sparse cluster formation or aggregation of disordered structures, could explain the experimental difficulties in obtaining crystals. The poor nucleation in conditions that otherwise allow growth is also in line with the recent suggestion that heterogeneous nucleation of protein crystals is likely to be dominant²⁸.

The degeneracy in constructing unit cells leads us to speculate that the high degree of freedom that the $P2_12_12_1$ space group offers could be both a strength and an obstruction for protein crystal growth. The strength resides in the adaptability of the unit cells, providing multiple options to allow macromolecules with no intrinsic symmetry to be incorporated into a periodic three-dimensional network¹⁸. The disadvantage on the other hand is that, although a small nucleus can be formed, these numerous possibilities could obstruct finding a suitable conformation that would grow into a full crystal.

ACKNOWLEDGMENTS

This work was supported by EPSRC/NSF World Materials Network grant EP/J018619/1 and EPSRC Programme Grant EP/I001352/1. We thank Abraham Lenhoff, Seth Fraden, Andela Šarić, Yasmine Chebaro,

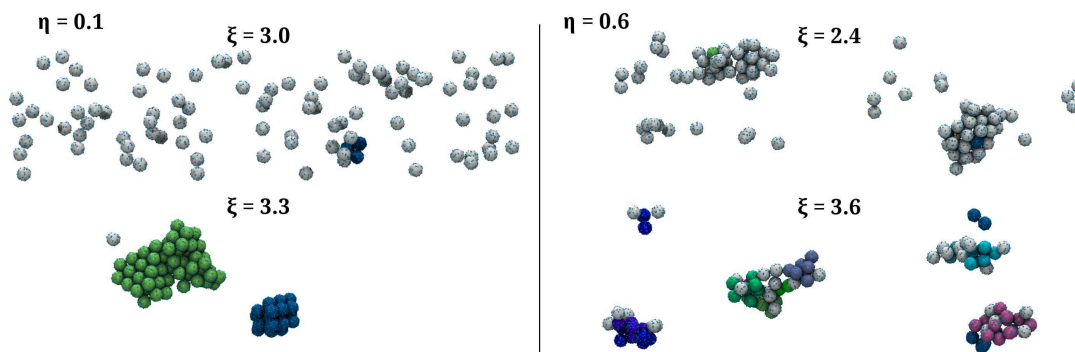


FIG. 10. **Crystal nuclei for different amounts of interfering interactions.** Snapshots from the end of four simulations. In the left panel $\eta = 0.1$ and $\xi = 3.0$ and 3.3 (upper and lower images, respectively). In the right panel $\eta = 0.6$ and $\xi = 2.4$ and 3.6 (upper and lower images, respectively). Particles colored gray have not created any crystalline bonds. Any other color indicates a crystalline cluster.

Julian Schrenk and Laura Filion for helpful discussions. Additional data related to this publication are available at the University of Cambridge data repository, <https://www.repository.cam.ac.uk/handle/1810/252594>.

- ¹A. Basu, C. P. Govardhan, C. W. Jung, and A. L. Margolin, *Expert Opin. Biol. Ther.* **4**, 301 (2004).
- ²S. R. McGuffee and A. H. Elcock, *J. Am. Chem. Soc.* **128**, 12098 (2006).
- ³P. Mereghetti, R. R. Gabdoulline, and R. C. Wade, *Biophys. J.* **99**, 3782 (2010).
- ⁴M. H. J. Hagen and D. Frenkel, *J. Chem. Phys.* **101**, 4093 (1994).
- ⁵A. Lomakin, N. Asherie, and G. B. Benedek, *J. Chem. Phys.* **104**, 1646 (1996).
- ⁶D. F. Rosenbaum and C. F. Zukoski, *J. Cryst. Growth* **169**, 752 (1996).
- ⁷D. F. Rosenbaum, A. Kulkarni, S. Ramakrishnan, and C. F. Zukoski, *J. Chem. Phys.* **111**, 9882 (1999).
- ⁸A. Lomakin, N. Asherie, and G. B. Benedek, *Proc. Natl. Acad. Sci. USA* **96**, 9465 (1999).
- ⁹R. Sear, *J. Chem. Phys.* **111**, 4800 (1999).
- ¹⁰M. Hloucha, J. F. M. Lodge, A. M. Lenhoff, and S. I. Sandler, *J. Cryst. Growth* **232**, 195 (2001).
- ¹¹H. Liu, S. K. Kumar, and F. Sciortino, *J. Chem. Phys.* **127**, 084902 (2007).
- ¹²D. Fusco and P. Charbonneau, *Phys. Rev. E* **88**, 012721 (2013).

- ¹³F. Roosen-Runge, F. Zhang, F. Schreiber, and R. Roth, *Sci. Rep.* **4**, 7016 (2014).
- ¹⁴I. Saika-Voivod, F. Romano, and F. Sciortino, *J. Chem. Phys.* **135**, 124506 (2011).
- ¹⁵D. Fusco and P. Charbonneau, *J. Phys. Chem. B* **118**, 8034 (2014).
- ¹⁶M. Kastelic, Y. V. Kalyuzhnyi, B. Hribar-Lee, K. A. Dill, and V. Vlachy, *Proc. Natl. Acad. Sci. USA* **112**, 6766 (2015).
- ¹⁷L. J. Quang, S. I. Sandler, and A. M. Lenhoff, *J. Chem. Theory Comput.* **10**, 835 (2014).
- ¹⁸S. W. Wukovitz and T. O. Yeates, *Nat Struct Biol* **2**, 1062 (1995).
- ¹⁹L. Filion, M. Marechal, B. van Oorschot, D. Pelt, F. Smalenburg, and M. Dijkstra, *Phys. Rev. Lett.* **103**, 188302 (2009).
- ²⁰H. Stokes and D. Hatch, *J. Appl. Cryst.* **38**, 237 (2005).
- ²¹See supplemental material at [URL will be inserted by AIP] for figures S1 (symmetry space groups in protein crystals), S2 (unit cell energies and particle volume fractions) and S3 (snapshots from simulations).
- ²²J. Janin and F. Rodier, *Proteins* **23**, 580 (1995).
- ²³A. George and W. W. Wilson, *Acta Crystallogr. D* **50**, 361 (1994).
- ²⁴N. E. Chayen, *Curr. Opin. Struct. Biol.* **14**, 577 (2004).
- ²⁵P. S. Sarangapani, S. D. Hudson, R. L. Jones, J. F. Douglas, and J. A. Pathak, *Biophys. J.* **108**, 724 (2015).
- ²⁶T. Haxton and S. Whitelam, *Soft Matter* **8**, 3558 (2012).
- ²⁷L. Hedges, R. Mannige, and S. Whitelam, *Soft Matter* **10**, 6404 (2014).
- ²⁸A. Akella, A. Mowitz, M. Heymann, and S. Fraden, *Cryst. Growth Des.* **14**, 4487 (2014).