Soft Matter Perspective on Protein Crystal Assembly

Diana Fusco^{c,b,a}, Patrick Charbonneau^b

^a*Departments of Physics and Integrative Biology, University of California, Berkeley, California 94720, USA*

^b*Department of Chemistry, Duke University, Durham, North Carolina 27708, USA*

^c*Program in Computational and Biology and Bioinformatics, Duke University, Durham, North Carolina 27708, USA*

Abstract

Crystallography may be the gold standard of protein structure determination, but obtaining the necessary high-quality crystals is also in some ways akin to prospecting for the precious metal. The tools and models developed in soft matter physics to understand colloidal assembly offer some insights into the problem of crystallizing proteins. This topical review describes the various analogies that have been made between proteins and colloids in that context. We highlight the explanatory power of patchy particle models, but also the challenges of providing guidance for crystallizing specific proteins. We conclude with a presentation of possible future research directions. This article is intended for soft matter scientists interested in protein crystallization as a self-assembly problem, and as an introduction to the pertinent physics literature for protein scientists more generally.

Keywords: protein phase diagram, patchy particle models, protein crystallization

1. Introduction

Biological macromolecules are central to life processes. Although some of these processes can be characterized at a relatively coarse scale, more often than not a microscopic understanding of the structure and dynamics of the involved

5 biomolecules is also essential¹. Proteins, for instance, interact with each other
 and their environment through fine-tuned structural features so as to perform
 their biological functions [1, 2, 3, 4, 5]; reciprocally, protein malfunction is often
 due to structural defects, and may result in diseases [6]. Proteins also represent
 some of the most sophisticated nano-machines known, having been shaped by
 10 natural selection. From an engineering perspective, few other systems offer more
 insights into designing devices on that size scale [7]. Our limited knowledge of
 protein structures thus limits our comprehension of biological phenomena, our
 ability to discover new drugs, and our design of bio-inspired materials [8, 9,
 10]. For that reason, sizable research efforts are being expended on extracting
 15 detailed protein structures and dynamics [11, 12].

As a field, structural biology mostly relies on protein structures obtained
 from diffraction-based methods. Since X-ray crystallography enabled the struc-
 ture determination of myoglobin and hemoglobin in the 50's, both the amount
 and quality of structural information have greatly increased [13]. Protein crys-
 20 tallography, however, requires protein crystals, whose obtention is often the lim-
 iting experimental step [14, 15]. Expertise from many different fields – from com-
 putational biology and robotics to surfactant science – has thus been brought to
 bear on the problem [8]. This article specifically reviews the contribution of soft
 matter physics to understanding the crystallization of globular proteins². Al-
 25 though the scientific conversation between the soft matter and structural biology
 communities has thus far mostly focused on providing a physical rationale for
 experimental observations, recent conceptual advances and an increased back
 and forth between theory and experiment suggest that richer exchanges may
 soon become the norm. In the following, we recapitulate how this development
 30 has come to be. We present in Sect. 2 and 3 an overview of the basics of protein

¹This phenomenon is sometimes playfully referred to as the revenge of structural biology.

²The assembly of membrane proteins typically relies on different physical processes [16]. These processes are generally beyond the scope of this review, although we get back to this issue in the conclusion.

crystallography and protein crystal assembly, respectively, in order to motivate the practical importance of physical insights. We then discuss various isotropic (Sect. 4) and patchy (Sect. 5) models of protein assembly. Section 6 concludes with a discussion of possible future research directions.

35 Note that we aim here to provide a reasonably broad physical description of protein crystallization and its constraints. Our target readership are soft matter scientists with either a novel or a renewed interest in the problem. We thus describe some more elementary aspects of the underlying structural biology and crystallization technology. We intend, however, the later sections to also
40 serve as an introduction to the soft matter literature for protein scientists more generally.

2. Diffraction and protein crystallization

Proteins are encoded in DNA as a series of base pairs, which are then translated into a sequence of amino acids, forming the primary structure of the
45 molecules. To perform its functions, a protein must also hierarchically fold into its secondary and tertiary structures. It is the properties of the resulting three-dimensional object that largely determine how the molecule interacts with its environment. This connection between function and shape is at the root of both structural biology and structural genomics.

50 Advances in high-throughput sequencing have markedly increased the number of protein-encoding genes and thus of known primary structures. Knowledge of the amino acid sequence alone, however, currently provides but limited insights into the higher-order structure of a protein. Predicting a protein's full tertiary structure from its sequence is indeed a remarkably complex task [17].
55 Because for most purposes protein structures cannot be inferred or calculated, they must be determined experimentally. The classical and most frequently used approaches for doing so are X-ray and neutron crystallography. These solid-phase diffraction methodologies can handle proteins containing several thousands of amino acids and reach sub-atomic resolution ($< 1.5 \text{ \AA}$). By contrast,

60 in spite of important methodologically advances, nuclear magnetic resonance
 (NMR) still cannot resolve the structure of solvated proteins containing more
 than two to three hundred amino acids, i.e., a few tens of kilodaltons [18, 19].
 NMR-resolved structures thus represent only a small (although increasing) frac-
 tion of the protein data bank (PDB). Recent advances have also pushed single-
 65 particle electron cryomicroscopy near atomic resolution [20, 21, 22]. However,
 (sub-)atomic resolution may remain physically unattainable because of radia-
 tion damage, beam-induced movement and charging of the sample [22]. The
 classical diffraction methodologies are thus likely to remain the reference for
 the foreseeable future. Interestingly, new diffraction-based techniques are also
 70 under development. Electrons, which scatter fairly strongly from molecules, can
 provide a diffracted image of a protein’s Coulomb potential [23], and the intense
 X-ray pulses from free-electron lasers (FEL) provide high-resolution structural
 information from relatively small crystals [24].

Whatever the chosen diffraction approach may be, three main steps must be
 75 performed in order to prepare a sample: (i) the protein must be expressed in a
 sufficient amount, which is typically achieved by using plasmids in cultured bac-
 terial cells; (ii) the protein must be purified, in order to isolate it from the biolog-
 ical medium used to express it; and (iii) the purified protein molecules must be
 assembled in an ordered and well-packed crystallite, whose minimal size depends
 80 on the scattering intensity of the diffracted radiation (from tens of nanometers
 or less on the side for FEL to millimeters for neutron beams [25, 26]). Although
 each of these steps presents several experimental challenges, crystallization is by
 far the most troublesome [8, 14, 9, 10]. Under standard biological conditions,
 most proteins do not easily crystallize. They likely evolved to limit aggregation
 85 that interferes with their normal biological function [27, 28] (with a few notable
 exceptions [8, 29]). Some neurodegenerative disorders, such as Alzheimer’s and
 Parkinson’s diseases, have indeed been linked to protein solubilization failure
 resulting in unwanted aggregation [30], and cataract formation can directly in-
 volve the crystallization of eye-lens proteins [31, 32, 33, 34]. Hence, in order to
 90 promote their periodic assembly, proteins must be placed in chemical conditions

that are typically far from those encountered in biological systems, yet must not result in a loss of secondary or ternary structure.

Even within those strict constraints, the chemical space within which to locate conditions that promote crystallization is remarkably vast. Exhaustive
95 searches are simply beyond experimental reach [35]. Crystallographers have thus developed chemical screens summarizing conditions that have been previously successful. For globular proteins, these are aqueous solutions containing various combinations of three families of co-solutes: inorganic salts, polymers (e.g., polyethylene glycol), and small organic molecules [8]. Technological advances
100 in automatizing the experimental process currently allow for up to thousands of these crystallization cocktails to be tested at once [36], but even that number is only a minute fraction of the full spectrum of chemical possibilities.

More crucially, existing screens are far from guaranteeing crystal formation. Despite accrued experience and improved experimental techniques, successful
105 crystallization indeed remains the exception rather than the rule. On average, only 0.04% of crystallization experiments generate good-quality crystals, which makes them quite time consuming and expensive to obtain [14]. Membrane proteins are even harder to crystallize [37]. Yet as long as diffraction-based methods remain the most desirable way to determine protein structures, a better
110 understanding of the physical mechanisms underlying protein crystal assembly is the only possible path towards a high-throughput scheme comparable to next-generation DNA sequencing [11, 38]. Brute-force approaches are unlikely to succeed on their own.

3. Solubility phase diagrams and protein crystallization

115 From a thermodynamic viewpoint, protein crystallization is a standard phase transition whereat the chemical potential of protein molecules dispersed in an aqueous solvent is equal to that of those ordered in a crystal. The transition is in some ways similar to water freezing into ice, suggesting that liquid-state descriptions may provide useful microscopic insights. Understanding protein

120 crystallization, or any other phase transition, should indeed follow from know-
ing the (effective) interactions between protein molecules [8, 39, 40, 41]. In
other words, a detailed description of protein-protein interactions should also
provide the protein phase behavior (Fig. 1) [42]. Yet, although the underlying
forces through which proteins interact, i.e., hydrogen bonding, van der Waals,
125 hydrophobic, and electrostatic interactions, are individually fairly well charac-
terized [43, 44, 45, 46, 47], their collective contribution to protein assembly is
much less well understood [48, 49]. A way to somehow coarsen microscopic
details into a simpler, effective description of protein pair interactions is thus
needed.

130 Phase diagrams capture graphically the regions of parameter space over
which different phases of matter are thermodynamically stable. For mixtures,
the phase diagram is multi-dimensional, but in some systems two-dimensional
projections on the temperature-concentration plane of a key component, e.g.,
the protein, recover a significant fraction of that information. For crystallization
135 cocktails, however, this projection typically involves rescaling the temperature
axis by an aggregate function of the solution conditions. Co-solutes can in-
deed change the energy scale over which protein molecules interact with one
another. For many proteins, this projection displays phases that are analogous
to those of a single-component system, for which the two-dimensional projection
140 is complete. Crystal, liquid, vapor, and supercritical fluid regimes are indeed
identifiable. The vapor phase corresponds to a low-concentration of proteins in
solution and the liquid to a high-concentration. The two phases are separated
by a first-order transition that terminates at a second-order critical point, above
which a supercritical fluid is observed. Because protein solutions are mixtures
145 of proteins, water, and other additives, however, the typical properties of phases
in a protein system may significantly differ from what is typically observed in
simple fluids. A protein crystal, for instance, contains on average more than
40% water [50] and may embed co-solutes in fractions that differ from what is
left in the crystallization cocktail.

150 Phase diagrams for about a dozen proteins, including lysozyme, γ -crystallin,

insulin, and myoglobin, have been experimentally determined under different solution conditions [51, 52, 53, 54, 55, 56, 57]. Intriguingly, these phase diagrams have in common a topology that is qualitatively different from that of simple liquids (Fig. 2) [58]. In simple liquids, the first-order phase transition between a vapor and a liquid phase terminates at a stable second-order critical point that lies *above* the crystal solubility curve, whereas in proteins the critical point is typically situated *below* the solubility line. The protein gas-liquid binodal and critical point are, therefore, metastable with respect to gas-crystal coexistence (Fig. 2).

It is also empirically found that successful crystallization typically is most common when a protein solution drop is prepared in the supersaturated region between the solubility line and the metastable critical point – a region that is sometimes called the nucleation zone or the crystallization gap (Fig. 2A) [8, 59, 60, 61]. If it is prepared at conditions below the critical point, the system has a propensity to aggregate in a disordered, percolating network, i.e., a gel; the gelation probability increases with quench depth [62, 63]. Such aggregation, although thermodynamically metastable with respect to crystal assembly, forms much more rapidly and is often long-lived. It thus reduces the likelihood of successful protein crystallization.

Based on this description, given a protein phase diagram, protein crystallization should be easily achieved (Fig 1). Unfortunately, experimental determination of protein phase diagrams (considering the vast number of possible co-solutes) is a lot more time- and resource-consuming than even the most ambitious crystallization screen. Hence, although physically elegant, a direct application of thermodynamics principles is of limited practical relevance. This picture is also overly reductive because a protein’s structure may also respond to changing solution conditions. In order to gain useful insights about protein crystal assembly, one must therefore extract the key features of their phase diagrams from a limited amount of information. Because atomistic models of crystal assembly would be computationally intractable (simulating a simple protein can be quite demanding, let alone hundreds of them), coarse-grained models are

Figure 1: Schematic description of the vicious circle of protein crystallization. Crystals are needed to obtain protein structures through diffraction (green arrow), but these crystals would be more easily obtained if experimental phase diagrams were known (black arrow). These phase diagrams are, however, expensive to determine. Phase diagrams of coarse-grained protein models may be accessible via simulations (blue arrow), but an appropriate protein model should stem from a detailed understanding of protein-protein interactions (red arrow), the computation of which requires some knowledge of the protein structure (purple arrow). By contextualizing the observed protein behavior, however, soft matter models could provide guidelines to refine and iterate the process.

the only viable option to fill in the missing information (Fig. 1). Conveniently, self-assembly has been heavily studied in related models over the last couple of decades.

185 4. Isotropic models of protein phase behavior

The field of soft matter physics has grown from a desire to understand the physics behind the assembly of squishy, mesoscale objects, such as polymers, liquid crystals, grains, and cells. From the self-assembly of these objects emerges a complex yet often universal array of material behaviors, including glass for-

190 mation and jamming, ordered microphases, and tissue growth. Soft materials are also particularly appealing because a large part of their complexity can be captured by purely classical models, and because the relevant range of these model's parameters is much wider than for atomic-scale simple liquids.

The study of the rich and robust phenomenology of colloidal suspensions, 195 which in some ways epitomizes the field of soft matter, opens the door to understanding protein assembly. In the mid-80's, suspensions of purely repulsive colloidal particles were first observed to crystallize similarly to classical hard spheres [64]. Adding a depletant, i.e., a soluble and inert co-solute much smaller than the colloids, was understood to result in a net pair attraction between particles 200 [65, 66], and, based on the law of corresponding states [58], was expected to result in the presence of a gas-liquid coexistence region. Experiments showed instead that adding a depletant unavoidably results in gel formation [67].

A possible resolution to this discrepancy emerged from the work of Lekkerkerker and Frenkel [68, 69], who found that decreasing the attraction range de- 205 presses the gas-liquid coexistence region more than it lowers the crystal solubility curve (Fig. 2B). For particles with a square-well attraction range $\lambda\sigma \lesssim 1.25\sigma$, relative to particles of diameter σ , the critical point thus falls below the solubility curve. Different models with short-range attraction confirmed the qualitative robustness of this result [70, 71, 72, 73, 74, 75, 76, 77], and its universality was 210 synthesized in an extended law of corresponding states [78, 79, 80].

The change in behavior in going from long- to short-range attraction can be intuitively explained by considering the energy-entropy balance in the liquid and crystal phases. For long-range attraction, there exists a relatively broad concentration-temperature range within which liquid particles are close enough 215 to benefit from each other's attraction while maintaining the high entropy characteristic of disordered configurations (Fig. 2B). The liquid free-energy can therefore be lower than that of the crystal, despite the crystal having a lower potential energy. When the attraction range is reduced, however, particles in the liquid have to be much closer to each other in order to feel their attraction 220 (Fig. 2B). This constraint drastically reduces the number of low-energy config-

urations, and thus the liquid entropy. By contrast, for interaction ranges as low as $\lambda\sigma = 1.10\sigma$, which is comparable to the interparticle distance in a hard-sphere crystal near melting [58], the crystal entropy remains almost unchanged. As a result, the liquid free-energy raises above that of the crystal and the liquid
225 phase ceases to be thermodynamically stable (Fig. 2B).

The gas-liquid coexistence, however, remains metastable; theoretical and experimental work has showed that its dynamical influence does not disappear. Homogeneous systems that are supersaturated near the metastable critical point indeed experience critical fluctuations that lower the barrier to crystal nucle-
230 ation [81]. In addition, if quenched below the critical point, these systems first undergo a spontaneous phase separation that is well described by spinodal decomposition. It is this process that gets arrested by the sluggish particle dynamics in the dense phase, and results in the emergence of a percolating gel network [82, 83, 84, 85, 86, 87].

This phase behavior is stunningly reminiscent of that of many proteins, as
235 described in Sect. 3. The analogy between short-range attractive colloids and proteins, which caught the attention of a number of soft matter scientists in the mid-90's [52, 88, 81, 89], is further supported by microscopic evidence. First, many globular proteins are roughly spherical objects with a fairly short-range
240 attraction. Their diameter is of at least few nanometers, while the mechanisms that result in pair attraction once electrostatic repulsion is screened, e.g., hydrogen bonding, salt bridges, and hydrophobicity, extend only over a few angstroms. These attractive forces are thus felt within about 10% of the protein diameter, which falls well within the specification for short-range attraction.
245 Second, George and Wilson observed that the optimal solution conditions for protein crystallization were consistent with being in proximity of a metastable critical point [90, 91, 92]. Third, contemporary studies of the PDB did not identify any statistical signatures of a preferred relative orientation of proteins chains in crystals, which was consistent with protein-protein attraction being
250 roughly orientation-independent, i.e., isotropic [93, 94]. This last suggestion, in particular, would prove to be overly simplistic, as we discuss in Sec. 5. A pro-

ductive effort at characterizing short-range isotropic models and subsequently rationalizing the phase diagram [95, 74] and solution behavior [96, 97, 98] of specific proteins as well as making generic arguments about heterogeneous nucleation [99, 100] from that perspective nonetheless ensued.

Figure 2: A: Typical topology of a temperature T -protein concentration phase diagram, where the effective interaction U can be tuned by changing solution conditions, e.g. changing salt concentration and pH, adding depletants, etc. At low saturation (yellow in the color bar), the drive to crystallize is insufficient to generate crystals within reasonable experimental time. At high saturation (red in the color bar), the drive to assemble is so strong that molecules form amorphous aggregates. In between lies the nucleation zone (or crystallization gap), where crystal assembly is possible (green in the color bar). By tuning T or U experimentally, the vertical axis can be rescaled and the saturation level controlled. B: The radial range of attraction, $\lambda\sigma$, of the pair interaction potential, $U(r)$, affects the topology of the scaled temperature-packing fraction, $T/\epsilon-\eta$, phase diagram. The situation is here illustrated for square-well fluids, but is fairly independent of the precise form of $U(r)$. In simple liquids, the relative interaction range is sufficiently large, $\lambda \gtrsim 1.25$, to make the liquid state entropically stable (top half). In colloids and proteins, the short-range attraction, i.e., for $\lambda \lesssim 1.25$, the reduced liquid entropy lowers the critical point below the solubility line (bottom half).

5. Patchy models of phase behavior

Though appealing in their simplicity, colloidal descriptions based on isotropic interactions between proteins miss some of the key phenomenology of globular proteins [101, 102, 52, 103, 55]. First, isotropic interactions systematically result in densely-packed crystals, such as the face-centered-cubic (FCC) or body-

centered-cubic (BCC) lattices (Fig. 3), while crystals of biomolecules are fairly *empty*. Their packing fraction η , i.e., the fraction of space physically occupied by proteins, is between 0.3 and 0.5 (compared to 0.74 and 0.68 for FCC and BCC, respectively), even though the protein chains themselves are essentially
265 close packed [104]. Second, the width of the region identified by the metastable vapor-liquid line in experimental protein phase diagrams does not match the expectations for particles with isotropic interactions [103, 105, 106]. Third, point mutations that change a single amino acid residue at the protein surface can dramatically alter the topology of a protein phase diagram [55, 107], while if
270 attraction were truly isotropic such a small perturbation would be expected to have but a fairly limited impact.

From a (bio)chemical viewpoint, these observations are far from surprising. Protein surfaces are heterogeneous and complex. There is thus no reason to believe that different relative protein orientations should result in similar inter-
275 actions. Unlike depletion attractions, hydrogen bonding and salt bridges are strongly directional and localized; they do not evenly cover the whole protein surface. Even the less chemically specific interactions based on hydrophobicity rely on the existence a particularly good structural match (recognition) between two parts of a protein surface, and thus exhibits a strong orientation dependence (see Appendix for a more detailed discussion of specificity). We also now
280 better appreciate, thanks to more careful statistical analysis of the PDB, that crystal contacts are quite different from randomly chosen elements of protein surfaces. They are enriched in glycine and small hydrophobic residues, and depleted in large polar residues with high side-chain entropy, such as lysine and
285 glutamic acids [108]. A large fraction of glycine and alanine on a protein surface further correlates with a higher probability of the molecule having been crystallized [109].

5.1. Simple Patches

This evidence drove the realization that directional aspects of protein pair
290 interactions ought to be taken qualitatively into account, in order to obtain rea-

Figure 3: Increasing the complexity of schematic models is necessary to accommodate the high interaction specificity that controls protein assembly. Isotropic interactions, which are well-suited to describe simple liquids or depletion interaction in colloids, exclusively depend on the inter-particle distance and assemble in close-packed crystals (left). Thermodynamic stability of low-density crystal requires directionality, which is introduced in patchy models by requiring the alignment of the surface patches for two particles to interact (middle, red patches have to face one another to interact). However, the highly heterogeneous protein surface requires that patches only interact with a single partner patch, and not promiscuously (right, only patches of the same color interact).

sonable minimal models of proteins (Fig. 3). Bonding directionality makes the solution behavior of proteins more akin to that of associative liquids, such as water, than to that of simple (isotropic) liquids, which is a fundamental distinction in liquid-state theory [58]. Nearly coincident with this realization, came the promise of synthetic colloids with an increasing degree of sophistication, such as faceted and DNA-coated particles, which led to a boon of interest in models with directional interactions, as was recently reviewed by Bianchi *et al.* [106]. Our aim here is not to broadly go over this advance, but to highlight the aspects most important to understanding protein crystallization.

A broad array of schematic models with short-range anisotropic, directional attraction, i.e., *patchy* particle models, have since been studied (Fig. 3). Although the details of patch geometry and interaction parameterization may differ, these patchy models generally have phase diagrams [110, 103, 111, 112, 106], fluid properties [113], and assembly pathways [114, 115, 116] that are signifi-

305 cantly different from those of short-range isotropic models. Most saliently for
 protein assembly, patchiness further lowers the metastable critical point, sta-
 bilizes open geometries such as diamond and cubic crystals, and qualitatively
 changes the shape of the gas-liquid coexistence region (Table 1) [117, 106]. This
 last point is interesting because experimental phase diagrams of proteins, such
 310 as lysozyme and γ -crystallin, have a binodal whose width and critical packing
 fraction are much smaller than those of isotropic models [57, 52, 103]. Although
 patchiness requires the specification of more model parameters and results in
 a certain loss of universality [118, 106], it also weakens high-order correlations
 in the fluid structure, which enables the use of relatively simple liquid-state de-
 315 scriptions, such as Wertheim’s theory, for their analysis [119, 120, 58, 121, 122,
 123, 124].

Because most of the theoretical studies of patchy particles were motivated
 by the promises of colloidal synthesis [125, 126], they often focused on particles
 with a couple of patches [117, 127, 128, 129], or on varying the surface coverage
 320 of a single patch, i.e., Janus particles [130, 131, 132]. For particles with one to
 three patches, small perturbations to the number of patches indeed dramatically
 affects the liquid behavior, notably enabling the stabilization of empty liquids –
 liquid states with a vanishing density [117, 133]. If this regime is of interest for
 protein crystal assembly, it is as one to be avoided. Monomeric protein crystals
 325 display at least six patches, which is a minimal for mechanical stability [94, 134].
 The lack of evolutionary pressure for proteins to crystallize indeed results in their
 crystals having most commonly the lowest-symmetry chiral (protein chains are
 chiral) point group compatible with that number of contacts, i.e., $P2_12_12_1$ [135].
 Crystals of oligomeric proteins, for which evolution has shaped at least the
 330 oligomeric contacts, often assemble in complex asymmetric unit cells with more
 than ten distinct patches [136, 137]. The competition between these interactions
 can favor the formation of small metastable crystallites, hindering the assembly
 of defectless crystals. Patchy models with a larger number of patches have,
 however, received limited attention thus far [138].

335 5.2. *Specific Patches*

In addition to being directional, protein-protein interactions are typically specific; there is a one-to-one match between pairs of crystal contacts (Fig. 3). Because pairwise attraction between proteins depends sensitively on molecular details, a given patch often interacts exclusively with a single other patch, to
340 the exclusion of all others.

Interaction specificity has a sizable effect on protein assembly. It obviously affects the liquid entropy and thus the position of the metastable critical point relative to the solubility curve [139, 140, 141, 142, 124], but more importantly it results in interaction heterogeneity. Because each pair of interacting
345 patches relies on different physicochemical mechanisms, their bonding strength varies [142]. Our recent study of bond energy asymmetry suggest that this factor may play a key role in protein crystallizability. Energy asymmetry alone can indeed result in gel formation due to percolation as in empty liquids, the closing of the crystallization gap, and the restabilization of the critical point above the
350 crystal solubility line [142]. Future studies will surely expand this list. Note that although no colloidal particle with specific patches as complex as those of proteins have yet been synthesized, patchy coatings of complementary selective DNA strands offer a synthetic gateway for obtaining such surface features [126].

5.3. *Measuring Patches*

355 The loss of generality that accompanies patchiness and specificity leads even schematic models of proteins to require a broad range of system-specific parameters. In order to develop a relevant model for a given protein, a better description of patch energetics and coverage is thus necessary. Although the PDB provides a wealth of structural information about crystal contacts [108, 143],
360 relating this information to an effective free energy of interaction in solution is not straightforward. Thus far, the energetics of very few crystal contacts has been characterized.

An early effort used a phenomenological model to estimate pair interactions from a PDB structure [139], in order to reconstruct the phase behavior of

Table 1: Summary of the strengths and weaknesses of short-range attractive models when applied to protein crystallization

Model	Strengths	Weaknesses
Isotropic	Minimal number of parameters	Incorrect crystal symmetry
	Metastable gas-liquid binodal	Incorrect binodal shape
		Robust to point mutations
Patchy	Few parameters	Unrealistic pair interactions
	Open crystals	Fixed geometry
	Reasonable bimodal shape	
Specific patches	Correct crystal symmetry	Many parameters
	Realistic pair interactions	Fixed geometry
	Tunable for specific proteins	

365 bovine chymotrypsinogen. Existing chemical databases do not suffice, however,
to generalize this approach with much accuracy. More recent attempts have
used all-atom molecular dynamics simulations of protein pairs in solutions in
order to extract the angularly-resolved potential of mean force [144, 124, 145].
Within the quality of the selected molecular force fields, these simulations offer
370 a reasonable characterization of known crystal contacts. Some of these studies
were even able to capture the crystal assembly behavior of various proteins. As
long as sufficient structural information about patchiness is available, crystal
patch energetics can thus be reconstructed reasonably well (Fig. 1).

For most systems, however, this information is not known a priori. It is thus
375 problematic to measure the patch characteristics of proteins that have not yet
been crystallized or that display patchy interactions that are incompatible with
the crystal structure. Because for the vast majority of proteins no structural
information is available at all, the former is a genuine difficulty. The latter
is also important, because evidence suggests that non-crystallographic contacts
380 may play a kinetic role in protein crystal assembly [146] and may even render
metastable crystal phases kinetically accessible [138].

Yet even if the structure of a protein of interest were fully known, brute

force molecular dynamics sampling of the relative surface of a pair of proteins is well beyond computational reach. A hierarchy of methods for identifying candidate attractive patches without complete exhaustion is a more promising way forward. Approaches that consider protein-protein “docking predictions” as well as structural homology, followed by testing of these suggestions by higher-precision methods have been used for that purpose [124], but it is still early days for the vicious circle of protein crystallization to be broken.

6. Conclusions and Open questions

Soft matter has thus far provided a qualitative physical perspective to the problem of protein crystal assembly. In order to push the analogy forward, both qualitative and quantitative improvements are needed.

First, as argued in Sec. 5.3, more and higher quality information about the protein interactions that give rise to patchiness must be obtained. For guiding the crystallization of protein homologues or for optimizing solution conditions so as to improve crystal quality, this information could be particularly useful. From a computational viewpoint, insights into angularly-resolved potentials of mean force could also be gleaned from large-scale simulations. If a sufficiently large database of interactions were available, it might even be conceivable to use statistical methods to parametrize the patchiness of protein-protein interactions without first extensively simulating the system. From an experimental viewpoint, careful studies of weak protein-protein interactions, as has recently been done for ubiquitin in solution would also be of much help [147]. Studying the structure of transient complexes in solution [148] may not only help identify potential crystal contacts, but also competing interactions that can hinder crystal assembly. The design of proteins that easily crystallize could also be used to validate and further enrich the microscopic insights obtained from the direct studies of protein-protein interactions [149].

Second, richer varieties of patchy models ought to be developed in order to address basic qualitative questions about protein crystal assembly.

1. What is the role of competing patches and dimer formation? Some proteins are observed to crystallize in more than one crystal lattice in the same solution composition [150, 151, 152, 153, 154, 155]. The type of crystal
415 that assembles can then depend on the initial protein concentration as well as on the experimental time and temperature.
2. What is the role of internal flexibility? The paradigm of a single well-folded protein structure is known to be overly simplistic [156], but it remains a key requirement for protein crystal assembly. Internal flexibility of the
420 protein may interfere with crystal assembly, but the physical constraints have not been given much attention.
3. What causes inverted solubility? Some proteins are characterized by a decreasing protein solubility with increasing temperature, i.e., an inverted solubility [157, 158, 159]. Sometimes a single mutation can flip the solubility
425 curve [55, 107] or significantly change the assembly kinetics [160, 14]. This phenomenon is tentatively attributed to the temperature dependence of the water entropy [161, 162], but remains poorly understood overall.
4. What are minimal models for membrane protein crystal assembly? As mentioned in the introduction, the crystallization of membrane proteins
430 typically involves assembly principles that are different than for globular proteins [163, 16]. Soft matter insights might be helpful in building better experimental guidance for this difficult, yet crucial [164], problem.

The coming years will thus likely see the emergence of increasingly rich patchy models that can provide a clearer physical understanding of these and related
435 processes.

In closing, it is important to note that patchy models have also found a use in the study of proteins beyond crystal assembly. The study of virus capsid formation, in particular, has greatly benefited [165]. The aggregation of proteins into amorphous structures, such as amyloids, is also within conceptual reach
440 of similar approaches [166, 167]. The absence of clear microscopic information on protein-protein contacts in these systems, however, provides an additional

challenge. The study of protein crystal assembly, which has the inherent advantage of providing structural feedback, when successful, may thus lead the way towards a better understanding of a broad array of protein-protein interactions.

445 **Acknowledgments**

We acknowledge the help of our colleagues and collaborators at Duke and beyond, who have patiently walked us through various aspects of protein science over the years. We also gratefully acknowledge support from National Science Foundation Grant no. NSF DMR-1055586.

450 **Appendix A. Characterizing crystal contacts: specific vs. non-specific protein-protein interactions**

Protein-protein interactions have often classified as being either *specific* and *non-specific*. Biological specificity is well-known to be a problematic label [168, Ch. 4], but it is nonetheless a quite prevalent characterization, including of
455 protein interactions that control crystal assembly. Its precise meaning, however, differs depending on the disciplinary context (Table A.2). In this appendix, we aim to identify these different significations and thus provide a brief guide to the reader of the scientific literature on the topic.

In chemistry, specificity distinguishes certain attraction forces from others,
460 although the classification of the various physical mechanisms is not unambiguous [43, (§ 18.8)]. In biophysics, the distinction between specific and non-specific interactions typically relies on the existence of an energy gap that clearly divides a single, strong (specific) interaction from the other (non-specific) ones [27, 169]. In molecular biology, specific interactions are deemed responsible
465 for the stoichiometric recognition of a given target, while non-specific interactions arise from the promiscuous yet non-biologically relevant association of molecules [93, 94, 170, 171]. Specific interactions have thus been evolutionarily tuned to be free-energetically strong and geometrically oriented, while non-specific attractions have not. As mentioned above, this general weakness, how-

Table A.2: Summary of the properties that differentiate specific from non-specific interactions in different fields of study

Field	Specific interactions	Non-specific interactions
Chemistry	Hydrogen bonds, salt bridges	Hydrophobic, depletion, van der Waals, electrostatic
Biophysics	Unique, strong, energetically gapped	Many and weak
Molecular biology	Evolutionary tuned, strong, geometrically constrained	Weak, randomly distributed

470 ever, may itself have evolved so as to prevent pathological aggregation [27, 28].
 Although these three definitions are not necessarily orthogonal to one another,
 we here specifically aim to clarify the last one.

When applied to crystal contacts, specificity has been used to suggest that
 these biologically non-functional interactions are in many ways indistinguishable
 475 from interfaces obtained by randomly bringing two proteins together [27]. These
 interfaces do present characteristics that are typical of non-specific interactions:
 they are weaker than functional interaction (of the order of few kJ/mol) and
 they do not have any obvious biological purpose. However, they also have unique
 properties that distinguish them from randomly selected surface patches [108,
 480 109]. Specific local protein properties are correlated with crystallization and
 protein surface regions carrying such properties are more likely to be responsible
 for the interactions that drive crystal formation. Hence, crystal contacts are
 triggered by the basic chemical interactions present in any molecular system,
 although they are typically different from the biologically functional ones. The
 485 importance of weak protein interactions is not limited to crystal contacts, but
 is also recognized in regard to the formation of transient protein complexes that
 are sometimes necessary for correct protein function [148].

References

- [1] T. L. Blundell, H. Jhoti, C. Abell, High-throughput crystallography for
490 lead discovery in drug design, *Nat. Rev. Drug. Discov.* 1 (1) (2002) 45–54.
- [2] P. Kuhn, K. Wilson, M. G. Patch, R. C. Stevens, The genesis of high-throughput structure-based drug discovery using protein crystallography, *Curr. Opin. Chem. Biol.* 6 (5) (2002) 704–710.
- [3] T. L. Blundell, S. Patel, High-throughput X-ray crystallography for drug
495 discovery, *Curr. Opin. Pharmacol.* 4 (5) (2004) 490–496.
- [4] I. Tickle, A. Sharff, M. Vinkovic, J. Yon, H. Jhoti, High-throughput protein crystallography and drug discovery, *Chem. Soc. Rev.* 33 (8) (2004) 558–565.
- [5] D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, Docking and scoring
500 in virtual screening for drug discovery: methods and applications, *Nat. Rev. Drug. Discov.* 3 (11) (2004) 935–949.
- [6] M. Congreve, C. W. Murray, T. L. Blundell, Structural biology and drug discovery, *Drug. Discov. Today* 10 (13) (2005) 895–907.
- [7] N. Huebsch, D. J. Mooney, Inspiration and application in the evolution of
505 biomaterials, *Nature* 462 (7272) (2009) 426–432.
- [8] A. McPherson, *Crystallization of Biological Macromolecules*, CSHL Press, Cold Spring Harbor, 1999.
- [9] N. E. Chayen, E. Saridakis, Protein crystallization: from purified protein to diffraction-quality crystal, *Nat. Meth.* 5 (2) (2008) 147–153.
- [10] N. E. Chayen, *High-Throughput Protein Crystallization*, Vol. 77, Academic Press, London, 2009, pp. 1–22.
510
- [11] M. Morange, The ambiguous place of structural biology in the historiography of molecular biology, in: H. Rheinberger, S. Chadarevian (Eds.),

- History and Epistemology of Molecular Biology and Beyond, Vol. 310,
 515 Max Planck Institute for the History of Science, 2006, pp. 179–188.
- [12] K. Khafizov, C. Madrid-Aliste, S. C. Almo, A. Fiser, Trends in structural coverage of the protein universe and the impact of the protein structure initiative, *Proc. Natl. Acad. Sci. U.S.A.* 111 (10) (2014) 3733–3738.
- [13] J. Nicola, Crystallography: Atomic secrets, *Nature* 505 (2014) 603–604.
- 520 [14] Q. Chen, P. G. Vekilov, R. L. Nagel, R. E. Hirsch, Liquid-liquid phase separation in hemoglobins: Distinct aggregation mechanisms of the $\beta 6$ mutants, *Biophys. J.* 86 (3) (2004) 1702–1712.
- [15] T. C. Terwilliger, D. Stuart, S. Yokoyama, Lessons from structural genomics, *Annu. Rev. Biophys.* 38 (1) (2009) 371–383.
- 525 [16] M. Caffrey, Crystallizing membrane proteins for structure determination: Use of lipidic mesophases, *Annu. Rev. Biophys.* 38 (1) (2009) 29–51.
- [17] K. A. Dill, J. L. MacCallum, The protein-folding problem, 50 years on, *Science* 338 (6110) (2012) 1042–1046.
- 530 [18] S. Grzesiek, H.-J. Sass, From biomolecular structure to functional understanding: new NMR developments narrow the gap, *Curr. Opin. Struct. Biol.* 19 (5) (2009) 585–595.
- [19] R. Kerfah, M. J. Plevin, R. Sounier, P. Gans, J. Boissbouvier, Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins, *Curr. Opin. Struct. Biol.* 32 (2015) 113 – 122.
- 535 [20] A. Doerr, Single-particle electron cryomicroscopy, *Nat. Meth.* 11 (1) (2014) 30–30.
- 540 [21] X. Li, P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard, Y. Cheng, Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM, *Nat. Meth.* 10 (6) (2013) 584–590.

- [22] X.-C. Bai, I. S. Fernandez, G. McMullan, S. H. Scheres, Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles, *eLife* 2 (2013) e00461.
- 545 [23] K. Yonekura, K. Kato, M. Ogasawara, M. Tomita, C. Toyoshima, Electron crystallography of ultrathin 3D protein crystals: Atomic model with charges, *Proc. Nat. Acad. Sci. U.S.A.* 112 (11) (2015) 3368–3373.
- [24] T. R. Barends, L. Foucar, S. Botha, R. B. Doak, R. L. Shoeman, K. Nass, J. E. Koglin, G. J. Williams, S. Boutet, M. Messerschmidt, I. Schlichting, De novo protein crystal structure determination from X-ray free-electron
550 laser data, *Nature* 505 (7482) (2014) 244–247.
- [25] H. P. Stevenson, A. M. Makhov, M. Calero, A. L. Edwards, O. B. Zeldin, I. I. Mathews, G. Lin, C. O. Barnes, H. Santamaria, T. M. Ross, S. M. Soltis, C. Khosla, V. Nagarajan, J. F. Conway, A. E. Cohen, G. Calero, Use of transmission electron microscopy to identify nanocrystals of chal-
555 lenging protein targets, *Proc. Natl. Acad. Sci. U.S.A.* 111 (23) (2014) 8470–8475.
- [26] M. P. Blakeley, P. Langan, N. Niimura, A. Podjarny, Neutron crystallography: opportunities, challenges, and limitations, *Curr. Opin. Struct. Biol.* 18 (5) (2008) 593–600.
- 560 [27] J. Janin, Protein-protein recognition, *Prog. Biophys. Mol. Biol.* 64 (2-3) (1995) 145–166.
- [28] J. P. K. Doye, A. A. Louis, M. Vendruscolo, Inhibition of protein crystallization by evolutionary negative design, *Phys. Biol.* 1 (2004) P9–P13.
- [29] J. P. K. Doye, W. C. K. Poon, Protein crystallization in vivo, *Curr. Opin. Colloid Interface Sci.* 11 (1) (2006) 40–46.
565
- [30] C. A. Ross, M. A. Poirier, Protein aggregation and neurodegenerative disease, *Nat. Med.* 10 (7) (2004) S10–S17.

- [31] S. Kmoch, J. Brynda, B. Asfaw, K. Bezouska, P. Novak, P. Rezacova, L. Ondrova, M. Filipec, J. Sedlacek, M. Elleder, Link between a novel human γ D-crystallin allele and a unique cataract phenotype explained by protein crystallography, *Hum. Mol. Gen.* 9 (12) (2000) 1779–1786.
- [32] A. Pande, J. Pande, N. Asherie, A. Lomakin, O. Ogun, J. King, G. B. Benedek, Crystal cataracts: human genetic cataract caused by protein crystallization, *Proc. Natl. Acad. Sci. U.S.A.* 98 (11) (2001) 6116–6120.
- [33] P. Evans, K. Wyatt, G. J. Wistow, O. A. Bateman, B. A. Wallace, C. Slingsby, The P23T cataract mutation causes loss of solubility of folded γ D-crystallin, *J. Mol. Biol.* 343 (2) (2004) 435–444.
- [34] F. F. Li, S. Z. Wang, C. Gao, S. G. Liu, B. J. Zhao, M. Zhang, S. Z. Huang, S. Q. Zhu, X. Ma, Mutation G61C in the CRYGD gene causing autosomal dominant congenital coralliform cataracts, *Mol. Vis.* 14 (46-53) (2008) 378–386.
- [35] J. Newman, E. E. Bolton, J. Müller-Dieckmann, V. J. Fazio, D. T. Gallagher, D. Lovell, J. R. Luft, T. S. Peat, D. Ratcliffe, R. A. Sayle, E. H. Snell, K. Taylor, P. Vallotton, S. Velanker, F. von Delft, On the need for an international effort to capture, share and use crystallization screening data, *Acta Cryst. F* 68 (3) (2012) 253–258.
- [36] A. E. Bruno, A. M. Ruby, J. R. Luft, T. D. Grant, J. Seetharaman, G. T. Montelione, J. F. Hunt, E. H. Snell, Comparing chemistry to outcome: The development of a chemical distance metric, coupled with clustering and hierarchal visualization applied to macromolecular crystallography, *PLoS ONE* 9 (6) (2014) e100782.
- [37] M. Caffrey, Membrane protein crystallization, *J. Struct. Biol.* 142 (1) (2003) 108–132.
- [38] H. Stevens, *Life Out of Sequence: A Data-Driven History of Bioinformatics*, University of Chicago Press, Chicago, 2013.

- [39] Z. Derewenda, Application of protein engineering to enhance crystallizability and improve crystal properties, *Acta Cryst. D* 66 (5) (2010) 604–615.
- [40] M. J. Anderson, C. L. Hansen, S. R. Quake, Phase knowledge enables rational screens for protein crystallization, *Proc. Natl. Acad. Sci. U.S.A.* 103 (45) (2006) 16746–16751.
- [41] E. Saridakis, N. E. Chayen, Towards a ‘universal’ nucleant for protein crystallization, *Trends Biotechnol.* 27 (2) (2009) 99–106.
- [42] A. C. Dumetz, A. M. Chockla, E. W. Kaler, A. M. Lenhoff, Comparative effects of salt, organic, and polymer precipitants on protein phase behavior and implications for vapor diffusion, *Cryst. Growth & Des.* 9 (2) (2009) 682–691.
- [43] J. N. Israelachvili, *Intermolecular and surface forces*, Academic Press, San Diego, 1991.
- [44] D. Chandler, Interfaces and the driving force of hydrophobic assembly, *Nature* 437 (7059) (2005) 640–647.
- [45] J. D. Gunton, A. Shiryayev, D. L. Pagan, *Protein Condensation*, Cambridge University Press, New York, 2007.
- [46] S. Granick, S. C. Bae, A curious antipathy for water, *Science* 322 (5907) (2008) 1477.
- [47] B. J. Berne, J. D. Weeks, R. Zhou, Dewetting and hydrophobic interaction in physical and biological systems, *Annu. Rev. Phys. Chem.* 60 (2009) 85–103.
- [48] P. G. Vekilov, A. A. Chernov, The physics of protein crystallization, *Solid State Phys.* 57 (2002) 1.

- [49] Y. D. Devedjiev, The role of flexibility and molecular shape in the crystallization of proteins by surface mutagenesis, *Acta Cryst. F* 71 (2) (2015) 157–162.
- [50] B. W. Matthews, Solvent content of protein crystals, *J. Mol. Biol.* 33 (2) (1968) 491–497.
- [51] M. Malfois, F. Bonnete, L. Belloni, A. Tardieu, A model of attractive interactions to account for fluid-fluid phase separation of protein solutions, *J. Chem. Phys.* 105 (8) (1996) 3290–3300.
- [52] A. Lomakin, N. Asherie, G. B. Benedek, Monte Carlo study of phase separation in aqueous protein solutions, *J. Chem. Phys.* 104 (4) (1996) 1646–1656.
- [53] M. Muschol, F. Rosenberger, Liquid–liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation/crystallization, *J. Chem. Phys.* 107 (6) (1997) 1953–1962.
- [54] A. Stradner, G. M. Thurston, P. Schurtenberger, Tuning short-range attractions in protein solutions: from attractive glasses to equilibrium clusters, *J. Phys.: Condens. Matter* 17 (31) (2005) S2805–S2816.
- [55] J. J. McManus, A. Lomakin, O. Ogun, A. Pande, M. Basan, J. Pande, G. B. Benedek, Altered phase diagram due to a single point mutation in human d-crystallin, *Proc. Natl. Acad. Sci. U.S.A.* 104 (43) (2007) 16856–16861.
- [56] S. Talreja, S. L. Perry, S. Guha, V. Bhamidi, C. F. Zukoski, P. J. A. Kenis, Determination of the phase diagram for soluble and membrane proteins, *J. Phys. Chem. B* 114 (13) (2010) 4432–4441.
- [57] C. Gögelein, D. Wagner, F. Cardinaux, G. Nägele, S. U. Egelhaaf, Effect of glycerol and dimethyl sulfoxide on the phase behavior of lysozyme: Theory and experiments, *J. Chem. Phys.* 136 (1) (2012) 015102.

- [58] J.-P. Hansen, I. R. McDonald, Theory of simple liquids, Academic Press, London, 2006.
- 650 [59] N. Asherie, Protein crystallization and phase diagrams, *Methods* 34 (3) (2004) 266–272.
- [60] N. Asherie, A dialogue about protein crystallization and phase diagrams, *Protein Pept. Lett.* 19 (7) (2012) 708–713.
- [61] M. Sleutel, J. F. Lutsko, D. Maes, A. E. S. Van Driessche, Mesoscopic
655 impurities expose a nucleation-limited regime of crystal growth, *Phys. Rev. Lett.* 114 (24) (2015) 245501.
- [62] J. A. Thomson, P. Schurtenberger, G. M. Thurston, G. B. Benedek, Binary liquid phase separation and critical phenomena in a protein/water solution, *Proc. Natl. Acad. Sci. U.S.A.* 84 (20) (1987) 7079–7083.
- 660 [63] A. C. Dumetz, A. M. Chockla, E. W. Kaler, A. M. Lenhoff, Protein phase behavior in aqueous solutions: Crystallization, liquid-liquid phase separation, gels, and aggregates, *Biophys. J.* 94 (2) (2008) 570–583.
- [64] P. N. Pusey, W. Vanmegen, Phase-behavior of concentrated suspensions of nearly hard colloidal spheres, *Nature* 320 (6060) (1986) 340–342.
- 665 [65] S. Asakura, F. Oosawa, On interaction between two bodies immersed in a solution of macromolecules, *J. Chem. Phys.* 22 (7) (1954) 1255–1256.
- [66] A. Vrij, Polymers at interfaces and interactions in colloidal dispersions, *Pure & Appl. Chem.* 48 (4) (1976) 471–483.
- [67] A. P. Gast, C. K. Hall, W. B. Russel, Polymer-induced phase separations
670 in non-aqueous colloidal suspensions, *J. Colloid Interface Sci.* 96 (1) (1983) 251–267.
- [68] H. N. W. Lekkerkerker, W. C. K. Poon, P. N. Pusey, A. Stroobants, P. B. Warren, Phase behaviour of colloid + polymer mixtures, *Europhys. Lett.* 20 (6) (1992) 559–564.

- 675 [69] M. H. J. Hagen, D. Frenkel, Determination of phase diagrams for the hard-core attractive Yukawa system, *J. Chem. Phys.* 101 (5) (1994) 4093–4097.
- [70] N. Asherie, A. Lomakin, G. B. Benedek, Phase diagram of colloidal solutions, *Phys. Rev. Lett.* 77 (23) (1996) 4832–4835.
- [71] F. W. Tavares, S. I. Sandler, Phase equilibria for the mean-force potential
680 of globular protein solutions, *AIChE J.* 43 (1) (1997) 218–231.
- [72] M. A. Miller, D. Frenkel, Competition of percolation and phase separation in a fluid of adhesive hard spheres, *Phys. Rev. Lett.* 90 (13) (2003) 135702.
- [73] H. Liu, S. Garde, S. Kumar, Direct determination of phase behavior of square-well fluids, *J. Chem. Phys.* 123 (17) (2005) 174505.
- 685 [74] D. L. Pagan, J. D. Gunton, Phase behavior of short-range square-well model, *J. Chem. Phys.* 122 (18) (2005) 184515.
- [75] R. López-Rendón, Y. Reyes, P. Orea, Thermodynamic properties of short-range square well fluid, *J. Chem. Phys.* 125 (8) (2006) 084508.
- [76] J. Largo, M. A. Miller, F. Sciortino, The vanishing limit of the square-well fluid: The adhesive hard-sphere model as a reference system, *J. Chem.*
690 *Phys.* 128 (13) (2008) 134513.
- [77] A. Fortini, E. Sanz, M. Dijkstra, Crystallization and gelation in colloidal systems with short-ranged attractive interactions, *Phys. Rev. E* 78 (4) (2008) 041402.
- 695 [78] M. G. Noro, D. Frenkel, Extended corresponding-states behavior for particles with variable range attractions, *J. Chem. Phys.* 113 (8) (2000) 2941–2944.
- [79] G. Foffi, F. Sciortino, Extended law of corresponding states in short-range square wells: A potential energy landscape study, *Phys. Rev. E* 74 (5)
700 (2006) 050401(R).

- [80] F. Platten, N. E. Valadez-Pérez, R. Castañeda-Priego, S. U. Egelhaaf, Extended law of corresponding states for protein solutions, *J. Chem. Phys.* 142 (17) (2015) 174905.
- [81] P. R. ten Wolde, D. Frenkel, Enhancement of protein crystal nucleation by critical density fluctuations, *Science* 277 (5334) (1997) 1975–1978.
- [82] G. Foffi, C. De Michele, F. Sciortino, P. Tartaglia, Arrested phase separation in a short-ranged attractive colloidal system: A numerical study, *J. Chem. Phys.* 122 (22) (2005) 224903.
- [83] S. Manley, H. M. Wyss, K. Miyazaki, J. C. Conrad, V. Trappe, L. J. Kaufman, D. R. Reichman, D. A. Weitz, Glasslike arrest in spinodal decomposition as a route to colloidal gelation, *Phys. Rev. Lett.* 95 (23) (2005) 238302.
- [84] S. Sastry, Journal club: A statistical physicist finds a way into a tangled web of matter, *Nature* 441 (7094) (2006) 671.
- [85] P. Charbonneau, D. R. Reichman, Phase behavior and far-from-equilibrium gelation in charged attractive colloids, *Phys. Rev. E* 75 (5) (2007) 050401(R).
- [86] S. Buzzaccaro, R. Rusconi, R. Piazza, “Sticky” hard spheres: Equation of state, phase diagram, and metastable gels, *Phys. Rev. Lett.* 99 (9) (2007) 098301.
- [87] P. J. Lu, E. Zaccarelli, F. Ciulla, A. B. Schofield, F. Sciortino, D. A. Weitz, Gelation of particles with short-range attraction, *Nature* 453 (7194) (2008) 499–503.
- [88] D. Rosenbaum, P. C. Zamora, C. F. Zukoski, Phase behavior of small attractive colloidal particles, *Phys. Rev. Lett.* 76 (1) (1996) 150–153.
- [89] N. E. Chayen, Methods for separating nucleation and growth in protein crystallisation, *Progr. Biophys. Mol. Biol.* 88 (2005) 329–337.

- [90] A. George, W. W. Wilson, Predicting protein crystallization from a dilute-solution property, *Acta Cryst. D* 50 (4) (1994) 361–365.
- 730 [91] C. Haas, J. Drenth, The protein-water phase diagram and the growth of protein crystals from aqueous solution, *J. Phys. Chem. B* 102 (21) (1998) 4226–4232.
- [92] W. W. Wilson, L. J. DeLucas, Applications of the second virial coefficient: protein crystallization and solubility, *Acta Cryst. F* 70 (5) (2014) 543–554.
- 735 [93] J. Janin, F. Rodier, Protein-protein interaction at crystal contacts, *Proteins* 23 (1995) 580–587.
- [94] O. Carugo, P. Argos, Protein-protein crystal-packing contacts, *Protein Sci.* 6 (10) (1997) 2261–2263.
- [95] G. Pellicane, D. Costa, C. Caccamo, Microscopic determination of the phase diagrams of lysozyme and γ -crystallin solutions, *J. Phys. Chem. B* 108 (23) (2004) 7538–7541.
- 740 [96] A. Stradner, G. Foffi, N. Dorsaz, G. Thurston, P. Schurtenberger, New insight into cataract formation: Enhanced stability through mutual attraction, *Phys. Rev. Lett.* 99 (19) (2007) 198103.
- 745 [97] N. Dorsaz, G. M. Thurston, A. Stradner, P. Schurtenberger, G. Foffi, Colloidal characterization and thermodynamic stability of binary eye lens protein mixtures, *J. Phys. Chem. B* 113 (6) (2009) 1693–1709.
- [98] G. Foffi, G. Savin, S. Bucciarelli, N. Dorsaz, G. M. Thurston, A. Stradner, P. Schurtenberger, Hard sphere-like glass transition in eye lens α -crystallin solutions, *Proc. Natl. Acad. Sci. U.S.A.* 111 (47) (2014) 16748–16753.
- 750 [99] N. E. Chayen, E. Saridakis, R. P. Sear, Experiment and theory for heterogeneous nucleation of protein crystals in a porous medium, *Proc. Natl. Acad. Sci. U.S.A.* 103 (3) (2006) 597–601.

- 755 [100] A. J. Page, R. P. Sear, Crystallization controlled by the geometry of a surface, *J. Am. Chem. Soc.* 131 (48) (2009) 17550–17551.
- [101] C. Haas, J. Drenth, W. W. Wilson, Relation between the solubility of proteins in aqueous solutions and the second virial coefficient of the solution, *J. Phys. Chem. B* 103 (14) (1999) 2808–2811.
- 760 [102] R. A. Curtis, H. W. Blanch, J. M. Prausnitz, Calculation of phase diagrams for aqueous protein solutions, *J. Phys. Chem. B* 105 (12) (2001) 2445–2452.
- [103] A. Lomakin, N. Asherie, G. B. Benedek, Aeolotopic interactions of globular proteins, *Proc. Natl. Acad. Sci. U.S.A.* 96 (1999) 9465–9468.
- 765 [104] F. M. Richards, The interpretation of protein structures: Total volume, group volume distributions and packing density, *J. Mol. Biol.* 82 (1) (1974) 1 – 14.
- [105] H. Liu, S. K. Kumara, F. Sciortino, Vapor-liquid coexistence of patchy models: Relevance to protein phase behavior, *J. Chem. Phys.* 127 (8) (2007) 084902.
- 770 [106] E. Bianchi, R. Blaak, C. N. Likos, Patchy colloids: state of the art and perspectives, *Phys. Chem. Chem. Phys.* 13 (14) (2011) 6397–410.
- [107] S. James, M. K. Quinn, J. J. McManus, The self assembly of proteins; probing patchy protein interactions, *Phys. Chem. Chem. Phys.* 17 (7) (2015) 5413–5420.
- 775 [108] M. Cieřlik, Z. S. Derewenda, The role of entropy and polarity in intermolecular contacts in protein crystals, *Acta Cryst. D* 65 (5) (2009) 500–509.
- 780 [109] W. N. Price, Y. Chen, S. K. Handelman, H. Neely, P. Manor, R. Karlin, R. Nair, J. F. Liu, M. Baran, J. Everett, S. C. N. Tong, F. Forouhar, S. S. Swaminathan, T. Acton, R. Xiao, J. R. Luft, A. Lauricella, G. T.

DeTitta, B. Rost, G. T. Montelione, J. F. Hunt, Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data, *Nat. Biotechnol.* 27 (1) (2009) 51–57.

- 785 [110] R. P. Sear, Phase behavior of a simple model of globular proteins, *J. Chem. Phys.* 111 (10) (1999) 4800–4806.
- [111] J. Chang, A. M. Lenhoff, S. I. Sandler, Determination of fluid-solid transitions in model protein solutions using the histogram reweighting method and expanded ensemble simulations, *J. Chem. Phys.* 120 (5) (2004) 3003–3014.
- 790 [112] C. Gögelein, G. Nägele, R. Tuinier, T. Gibaud, A. Stradner, P. Schurtenberger, A simple patchy colloid model for the phase behavior of lysozyme dispersions, *J. Chem. Phys.* 129 (8) (2008) 085102.
- [113] B. L. Neal, D. Asthagiri, A. M. Lenhoff, Molecular origins of osmotic second virial coefficients of proteins, *Biophys. J.* 75 (5) (1998) 2469–2477.
- 795 [114] H. Liu, S. K. Kumar, J. F. Douglas, Self-assembly-induced protein crystallization, *Phys. Rev. Lett.* 103 (2009) 018101.
- [115] S. Whitlam, Control of pathways and yields of protein crystallization through the interplay of nonspecific and specific attractions, *Phys. Rev. Lett.* 105 (8) (2010) 088102.
- 800 [116] T. K. Haxton, S. Whitlam, Design rules for the self-assembly of a protein crystal, *Soft Matter* 8 (13) (2012) 3558–3562.
- [117] E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli, F. Sciortino, Phase diagram of patchy colloids: Towards empty liquids, *Phys. Rev. Lett.* 97 (16) (2006) 168301.
- 805 [118] G. Foffi, F. Sciortino, On the possibility of extending the Noro-Frenkel generalized law of correspondent states to nonisotropic patchy interactions, *J. Phys. Chem. B* 111 (33) (2007) 9702–9705.

- [119] M. S. Wertheim, Fluids with highly directional attractive forces. I. Statistical thermodynamics, *J. Stat. Phys.* 35 (1-2) (1984) 19–34.
- 810 [120] M. S. Wertheim, Fluids with highly directional attractive forces. II. Thermodynamic perturbation theory and integral equations, *J. Stat. Phys.* 35 (1-2) (1984) 35–47.
- [121] E. Bianchi, P. Tartaglia, E. La Nave, F. Sciortino, Fully solvable equilibrium self-assembly process: fine-tuning the clusters size and the connectivity in patchy particle systems, *J. Phys. Chem. B* 111 (40) (2007) 11765–11769.
- 815 [122] E. Bianchi, P. Tartaglia, E. Zaccarelli, F. Sciortino, Theoretical and numerical study of the phase diagram of patchy colloids: Ordered and disordered patch arrangements, *J. Chem. Phys.* 128 (14) (2008) 144504.
- 820 [123] E. Bianchi, R. Blaak, C. N. Likos, Patchy colloids: state of the art and perspectives, *Phys. Chem. Chem. Phys.* 13 (14) (2011) 6397–6410.
- [124] D. Fusco, J. J. Headd, A. De Simone, J. Wang, P. Charbonneau, Characterizing protein crystal contacts and their role in crystallization: rubredoxin as a case study, *Soft Matter* 10 (2) (2014) 290–302.
- 825 [125] S. C. Glotzer, M. J. Solomon, Anisotropy of building blocks and their assembly into complex structures, *Nat. Mater.* 6 (7) (2007) 557–562.
- [126] Y. Wang, Y. Wang, D. R. Breed, V. N. Manoharan, L. Feng, A. D. Hollingsworth, M. Weck, D. J. Pine, Colloids with valence and specific directional bonding, *Nature* 491 (7422) (2012) 51–55.
- 830 [127] P. Charbonneau, D. Frenkel, Gas-solid coexistence of adhesive spheres, *J. Chem. Phys.* 126 (19) (2007) 196101.
- [128] F. Romano, E. Sanz, F. Sciortino, Phase diagram of a tetrahedral patchy particle model for different interaction ranges, *J. Chem. Phys.* 132 (18) (2010) 184501.

- 835 [129] F. Romano, E. Sanz, F. Sciortino, Crystallization of tetrahedral patchy particles in silico, *J. Chem. Phys.* 134 (17) (2011) 174502.
- [130] R. Fantoni, D. Gazzillo, A. Giacometti, M. A. Miller, G. Pastore, Patchy sticky hard spheres: analytical study and Monte Carlo simulations, *J. Chem. Phys.* 127 (23) (2007) 234507.
- 840 [131] S. Granick, S. Jiang, Q. Chen, Janus particles, *Physics Today* 2009 (July) (2009) 68–69.
- [132] G. Munaò, Z. Preisler, T. Vissers, F. Smalenburg, F. Sciortino, Cluster formation in one-patch colloids: low coverage results, *Soft Matter* 9 (9) (2013) 2652–2661.
- 845 [133] B. Ruzicka, E. Zaccarelli, L. Zulian, R. Angelini, M. Sztucki, A. Moussaid, T. Narayanan, F. Sciortino, Observation of empty liquids and equilibrium gels in a colloidal clay, *Nat. Mat.* 10 (1) (2010) 56–60.
- [134] E. Krissinel, Crystal contacts as nature’s docking solutions, *J. Comput. Chem.* 31 (1) (2010) 133–143.
- 850 [135] S. W. Wukovitz, T. O. Yeates, Why protein crystals favor some space-groups over others, *Nat. Struct. Biol.* 2 (12) (1995) 1062–1067.
- [136] S. Dasgupta, G. H. Iyer, S. H. Bryant, C. E. Lawrence, J. A. Bell, Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers, *Proteins: Structure, Function, and Bioinformatics* 28 (4) (1997) 494–514.
- 855 [137] M. Chruszcz, W. Potrzebowski, M. D. Zimmerman, M. Grabowski, H. Zheng, P. Lasota, W. Minor, Analysis of solvent content and oligomeric states in protein crystals—does symmetry matter?, *Prot. Sci.* 17 (4) (2008) 623–632.
- 860 [138] D. Fusco, P. Charbonneau, Competition between monomeric and dimeric crystals in schematic models for globular proteins, *J. Phys. Chem. B* 118 (28) (2014) 8034–8041.

- [139] M. Hloucha, J. F. M. Lodge, A. M. Lenhoff, S. I. Sandler, A patch-antipatch representation of specific protein interactions, *J. Cryst. Growth* 232 (1-4) (2001) 195–203.
- [140] N. M. Dixit, C. F. Zukoski, Crystal nucleation rates for particles experiencing anisotropic interactions, *J. Chem. Phys.* 117 (18) (2002) 8540–8550.
- [141] N. Dorsaz, L. Fillion, F. Smalenburg, D. Frenkel, Spiers memorial lecture: Effect of interaction specificity on the phase behaviour of patchy particles, *Farad. Disc.* 159 (2012) 9–21.
- [142] D. Fusco, P. Charbonneau, Crystallization of asymmetric patchy models for globular proteins in solution, *Phys. Rev. E* 88 (1) (2013) 012721.
- [143] N. W. Price II, S. K. Handelman, J. K. Everett, S. N. Tong, A. Bracic, J. D. Luff, V. Naumov, T. Acton, P. Manor, R. Xiao, B. Rost, G. T. Montelione, J. F. Hunt, Large-scale experimental studies show unexpected amino acid effects on protein expression and solubility in vivo in *E. coli*, *Microb. Inform. Exp.* 1 (1) (2011) 1–20.
- [144] G. Pellicane, G. Smith, L. Sarkisov, Molecular dynamics characterization of protein crystal contacts in aqueous solutions, *Phys. Rev. Lett.* 101 (24) (2008) 248102.
- [145] A. Taudt, A. Arnold, J. Pleiss, Simulation of protein association: Kinetic pathways towards crystal contacts, *Phys. Rev. E* 91 (3) (2015) 033311.
- [146] J. D. Schmit, K. Dill, Growth rates of protein crystals, *J. Am. Chem. Soc.* 134 (9) (2012) 3934–3937.
- [147] Z. Liu, W.-P. Zhang, Q. Xing, X. Ren, M. Liu, C. Tang, Noncovalent dimerization of ubiquitin, *Angew. Chem. Intl. Ed.* 51 (2) (2012) 469–472.
- [148] S. E. Acuner Ozbabacan, H. B. Engin, A. Gursoy, O. Keskin, Transient protein-protein interactions, *Protein Eng. Des. Sel* 24 (9) (2011) 635–648.

- [149] C. J. Lanci, C. M. MacDermaid, S.-g. Kang, R. Acharya, B. North,
890 X. Yang, X. J. Qiu, W. F. DeGrado, J. G. Saven, Computational design of a protein crystal, *Proc. Natl. Acad. Sci. U.S.A.* 109 (19) (2012) 7304–7309.
- [150] M. V. King, B. S. Magdoff, M. B. Adelman, D. Harker, Crystalline forms
of bovine pancreatic ribonuclease: techniques of preparation, unit cells,
895 and space groups, *Acta Cryst.* 9 (5) (1956) 460–465.
- [151] M. V. King, J. Bello, E. H. Pignataro, D. Harker, Crystalline forms
of bovine pancreatic ribonuclease. Some new modifications, *Acta Cryst.*
15 (2) (1962) 144–147.
- [152] M. Dixon, H. Nicholson, L. Shewchuk, W. Baase, B. Matthews, Structure
900 of a hinge-bending bacteriophage T4 lysozyme mutant, Ile3 $\rightarrow$ Pro, *J. Mol. Biol.* 227 (3) (1992) 917–933.
- [153] H. Faber, B. Matthews, A mutant T4 lysozyme displays five different
crystal conformations, *Nature* 348 (6298) (1990) 263–266.
- [154] D. E. McRee, S. M. Redford, T. E. Meyer, M. A. Cusanovich, Crystal-
905 lization and characterization of *Chromatium vinosum* Cytochrome *c'*, *J. Biol. Chem.* 265 (10) (1990) 5364–5365.
- [155] A. V. Elgersma, M. Ataka, T. Katsura, Kinetic studies on the growth of
three crystal forms of lysozyme based on the measurement of protein and
Cl⁻ concentration changes, *J. Cryst. Growth* 122 (1-4) (1992) 31–40.
- 910 [156] P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular
signaling and regulation, *Nature Rev. Mol. Cell. Biol.* 16 (1) (2015) 18–
29.
- [157] M. Budayova-Spano, F. Bonnet, J.-P. Astier, S. Veessler, Investigation of
aprotinin (BPTI) solutions during nucleation, *J. Cryst. Growth* 235 (1-4)
915 (2002) 547–554.

- [158] M. Budayova-Spano, S. Lafont, J.-P. Astier, C. Ebel, S. Veessler, Comparison of solubility and interactions of aprotinin (BPTI) solutions in H₂O and D₂O, *J. Cryst. Growth* 217 (3) (2000) 311–319.
- 920 [159] S. Veessler, N. Fert, M.-S. Costes, M. Czjzek, J.-P. Astier, Temperature and pH effect on the polymorphism of aprotinin BPTI in sodium bromide solutions, *Cryst. Growth & Design* 4 (6) (2004) 1137–1141.
- [160] N. Asherie, J. Pande, A. Pande, J. A. Zarutskie, J. Lomakin, A. Lomakin, O. Ogun, L. J. Stern, J. King, G. B. Benedek, Enhanced crystallization of the Cys18 to Ser mutant of bovine γ B crystallin, *J. Mol. Biol.* 314 (4) 925 (2001) 663–669.
- [161] A. Shirayev, D. L. Pagan, J. D. Gunton, D. S. Rhen, A. Saxena, T. Lookman, Role of solvent for globular proteins in solution, *J. Chem. Phys.* 122 (23) (2005) 234911.
- [162] N. Wentzel, J. D. Gunton, Effect of solvent on the phase diagram of a 930 simple anisotropic model of globular proteins, *J. Phys. Chem. B* 112 (26) (2008) 7803–7809.
- [163] E. P. Carpenter, K. Beis, A. D. Cameron, S. Iwata, Overcoming the challenges of membrane protein crystallography, *Curr. Opin. Struct. Biol.* 18 (5) (2008) 581–586.
- 935 [164] A. Bhattacharya, Protein structures: Structures of desire, *Nature* 459 (7243) (2009) 24.
- [165] M. F. Hagan, Modeling viral capsid assembly, *Adv. Chem. Phys.* 155 (2014) 1–68.
- 940 [166] N. S. Bieler, T. P. J. Knowles, D. Frenkel, R. Vácha, Connecting macroscopic observables and microscopic assembly events in amyloid formation using coarse grained simulations, *PLoS Comput. Biol.* 8 (10) (2012) e1002692.

- [167] A. Šarić, Y. C. Chebaro, T. P. J. Knowles, D. Frenkel, Crucial role of non-specific interactions in amyloid nucleation, *Proc. Natl. Acad. Sci. U.S.A.* 111 (50) (2014) 17869–17874.
- [168] L. E. Kay, *Who Wrote the Book of Life?: A History of the Genetic Code*, Stanford University Press, Stanford, 2000.
- [169] M. E. Johnson, G. Hummer, Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks, *Proc. Natl. Acad. Sci. U.S.A.* 108 (2) (2011) 603–608.
- [170] K. D. Wilkinson, Quantitative analysis of protein-protein interactions, in: *Protein-Protein Interactions: Methods and Protocols*, Vol. 261, Humana Press, New York, 2004, pp. 15–31.
- [171] T. Zhuang, B. K. Jap, C. R. Sanders, Solution NMR approaches for establishing specificity of weak heterodimerization of membrane proteins, *J. Am. Chem. Soc.* 133 (50) (2011) 20571–20580.