The physics of protein self-assembly

Jennifer J. McManus [a], Patrick Charbonneau [b], Emanuela Zaccarelli [c,d], Neer Asherie [e,f,*]

[a] Department of Chemistry, Maynooth University, Maynooth, Co. Kildare, Ireland
[b] Department of Chemistry, Duke University, Durham, NC, USA
[c] CNR ISC, UOS Sapienza, Piazzale A. Moro 5, 00185 Rome, Italy
[d] Department of Physics, Sapienza University of Rome, Piazzale A. Moro 5, 00185 Rome, Italy
[e] Department of Physics, Yeshiva University, New York, NY, USA
[f] Department of Biology, Yeshiva University, New York, NY, USA

ABSTRACT

Understanding protein self-assembly is important for many biological and industrial processes. Proteins can self-assemble into crystals, filaments, gels, and other amorphous aggregates. The final forms include virus capsids and condensed phases associated with diseases such as amyloid fibrils. Although seemingly different, these assemblies all originate from fundamental protein interactions and are driven by similar thermodynamic and kinetic factors. Here we review recent advances in understanding protein self-assembly through a soft condensed matter perspective with an emphasis on three specific systems: globular proteins, viruses, and amyloid fibrils. We conclude with a discussion of unanswered questions in the field.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The self-assembly of proteins into small-scale complexes plays a crucial biological role [1]. Under certain conditions, proteins also self-assemble into various structures that range from nm to μm in size (Fig. 1). This process is almost as ubiquitous as complexation and is equally essential to biology. Some proteins, such as those that make up viral capsids or the outer shell of bacterial microcompartments, self-assemble by design [2••,3••]. Others do so when something goes wrong: a conformational change triggers the aggregation of amyloid β-protein (Aβ) into fibrils [4] and a single-point mutation in hemoglobin (Hb) leads to its polymerization [5••]. This type of assembly can also result from simple changes to solution conditions (pH, temperature, ionic strength, cosolutes, etc.) [6].

Understanding protein self-assembly is fundamental to many physiological and industrial processes. For example, the fibrillization of Aβ is a feature of Alzheimer's disease [7] and the polymerization of mutant Hb is the primary pathogenic event in sickle-cell anemia [5••]; other protein condensation diseases, for which the pathology is associated with the self-assembly of a condensed protein phase, include

cataract formation [8] and Parkinson's disease [9]. In the industrial production of proteins, self-assembly can be harnessed for protein purification through crystallization or liquid–liquid phase separation [10], or can be problematic if encountered during formulation and storage (often at high protein concentration), when the assembly process is not controlled [11]. Protein self-assembly is also essential to structural biology. Most structures are determined through x-ray crystallography, which requires the production of high-quality protein crystals [12••].

Here we review recent advances in understanding protein self-assembly. We adopt a soft condensed matter perspective in which simplified models are used to capture the essential elements of protein interactions to determine their assembly. While it is true that atomic-level details are sometimes required for a complete explanation of specific phenomena, the near ubiquitous nature of protein self-assembly suggests the existence of universal elements governing it, which many experimental, computational, and theoretical findings support. We focus on three specific systems: globular proteins, viral capsids, and amyloid fibrils. These systems formed the core of a 2015 CECAM workshop we organized that brought together researchers from diverse fields (including material science, crystallography, macromolecular chemistry, and biophysics) to discuss current challenges in understanding protein self-assembly. This opinion piece builds upon the presentations and discussions at the workshop as well as our own work in the field to stimulate further research—and perhaps breakthroughs—in the physics of protein self-assembly.

* Corresponding author at: Department of Physics, Yeshiva University, New York, NY, USA.
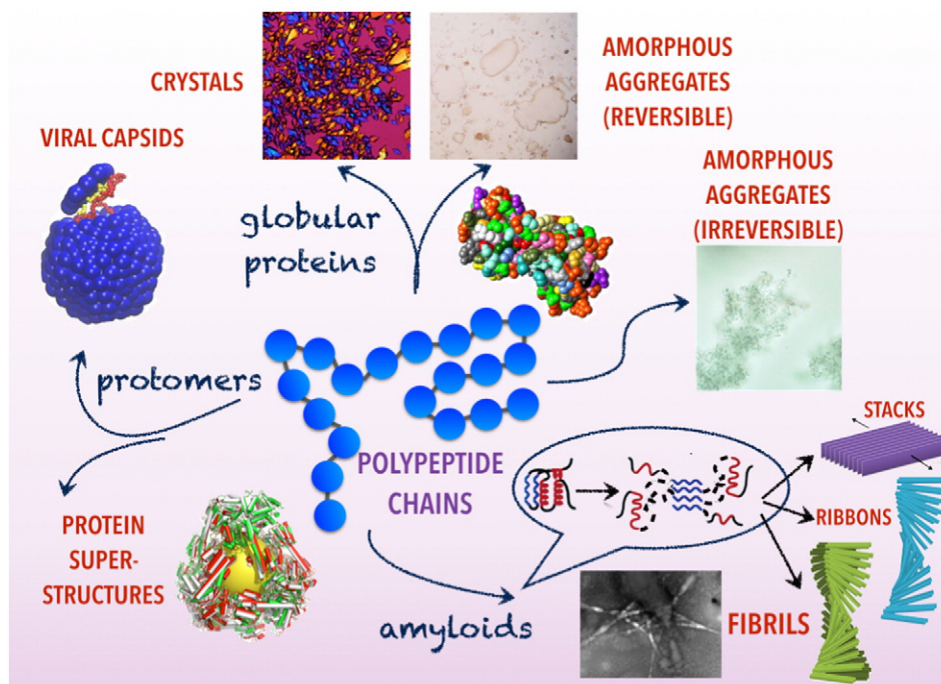E-mail address: asherie@yu.edu (N. Asherie).

Fig. 1. A number of different assemblies can be formed by proteins or peptides as illustrated here. Globular proteins can assemble to become either crystals or amorphous aggregates; both types of assembly are reversible (top). Upon a conformational change to a protein or peptide, either amorphous aggregates (which are irreversible) or amyloid fibrils can form (right). For some peptides, a range of different twisted structures have been observed including ribbons, fibrils and stacks (bottom right). Higher-order assemblies including viral capsids (in the presence of a nucleic acid) and protein superstructures (including bacterial microcompartments) can also form either naturally or in a directed manner, e.g. by mutagenesis (left).

2. Globular proteins

Globular proteins have a compact, often spherical shape with most of hydrophobic residues buried in the interior, and polar or charged residues predominantly at the surface, in contact with the solvent. Single domain globular proteins range in size from 6 to 300 kDa (roughly 60–2500 amino acid residues), and multiple domain proteins can be even larger. Globular proteins self-assemble into a variety of states: crystals, dense liquid phases, gels, fibers, and amorphous aggregates [13•]. A given globular protein can self-assemble into different states (e.g., crystal vs. aggregate) depending on solution conditions [14] and in some cases, two different states (e.g., two crystal forms) can even coexist [15–17].

The range of self-assembled states can be summarized in a phase diagram, and for several globular proteins, such as lysozyme [18], γ-crystallins [19], and bovine pancreatic trypsin inhibitor [20], comprehensive phase diagrams have been determined. These phase diagrams share a few common features, the most prominent being that liquid–liquid phase separation is metastable with respect to crystallization (Fig. 2). This metastability can be explained by modeling the globular protein as a simple attractive colloid: a hard spherical core with an isotropic attractive interaction [21]. Numerous theoretical and computational studies have confirmed that for sufficiently short-range attraction (less than about one-quarter the radius of the hard core), liquid–liquid phase separation becomes metastable with respect to crystallization [22], but arrest within the spinodal region of the phase diagram can also occur [23].

The isotropic model does not, however, provide a complete explanation of the self-assembly of globular proteins [24•]. Several predictions of isotropic models, such as the shape of the phase boundaries or the crystal density, differ from the experimental observations [25••]. While it is possible to modify the isotropic model so that its predictions agree more closely with experiments (e.g., introducing a temperature-dependent energy of interaction or two different ranges of interactions), marked discrepancies remain.

A natural way to improve upon isotropic models is to introduce anisotropy in the protein–protein attraction [26•,27•]. This naturally follows from the interaction between surface residues (amino acids)
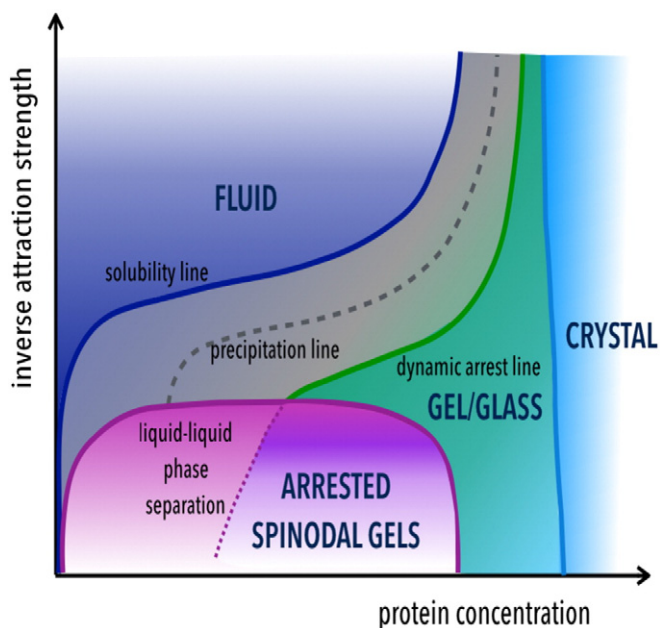

Fig. 2. Illustration of a state diagram for a globular protein indicating the variety of condensed phases it can form. Not all states are present for any given protein and the positions of the boundaries depend on solution conditions. The position of the precipitation line can occur anywhere in the grey region.

differing strongly from one to another [28]. (To a large extent, this effect dominates over shape anisotropy for simple globular proteins—see Conclusions.) A key challenge related to anisotropic models that has yet to be met in a comprehensive fashion is selecting the degree of anisotropy. A common approach consists of placing patches (either randomly or symmetrically) on the surface of a hard sphere and of imposing rules for the interaction between different patches (range, angular width, strength, etc.) [29•,30•]. By investigating models with different parameters (including a different number of patches), it is often possible to fit experimental results at hand [31]. While this approach can be effective, it is computationally intensive and must be repeated for each protein and corresponding set of experimental data. It also cannot identify a unique anisotropic model. As a result, it can be challenging to obtain both specific guidance for a given protein and general insights applicable to broad classes of proteins.

Another approach is to use a subset of the experimental results, typically a crystal structure, to determine a specific anisotropic model ("patchiness") and then use this model to analyze other self-assembly data [32]. This approach has not been explored as extensively as the first one, and more work should be carried out in this direction to assess its applicability to many different proteins. A significant limitation is its reliance on the crystal structure. While this information provides information about the patchiness of the protein, it limits the approach to those proteins that have already been crystallized.

Given the importance of protein crystals in structural biology, we believe that one long-term goal should be to develop models for protein self-assembly that help to directly predict crystal and possibly other forms of assembly. At present, protein crystallization is a low-success, brute-force endeavor in which solution conditions are changed or other alterations to the protein interactions are made (e.g., reduction of surface entropy by substituting floppy amino acids with more compact ones) in the hope that crystals are produced [12]. Although various proposals to enhance the likelihood that a protein crystallizes based on our understanding of self-assembly have been made, existing tools remain relatively crude [26•,33].

One notable tool is the Surface Entropy Reduction prediction (SERp) server that, given a protein sequence, suggests sites at which to make mutations that would reduce the surface entropy of the protein and therefore increase the likelihood that the protein analog crystallizes [34•]. It would, however, be useful to have a complementary tool that makes specific predictions regarding crystal self-assembly based on protein interactions. A first step towards this goal would be to devise a method for determining the patchiness of a protein that goes beyond random or symmetric guesses, yet does not require crystallization data a priori. One way to do so would be to identify regions based on the protein sequence and predictions of the protein structure that are likely to be crystal contacts. These regions would then be modeled as attractive or repulsive patches whose properties are related to the underlying molecular interactions of the regions—a charged patch would be modeled differently from a polar one. A tentative phase diagram could then be calculated for different solution conditions and those that lead to stable crystals forms identified and tested experimentally.

3. Viral capsids

A virus consists of a single copy of a genome (RNA or DNA) inside a closed, protein shell, known as a capsid that protects the nucleic acids. For viruses that encapsulate RNA, capsid self-assembly can occur spontaneously upon mixing protein and nucleic acid, which makes them ideal model systems to study the assembly process itself [35]. The capsid of these viruses is also relatively simple: a single protein molecule thick and often icosahedral (although other polyhedral and helical shapes are also observed) [36]. For example, the 28 nm capsid of the cowpea chlorotic mottle virus (CCMV) consists of 180 copies of a single capsid protein (of molecular weight ~20 kDa) [37].

As with the assembly of globular proteins, viral capsid assembly is sensitive to solution conditions, in particular pH and ionic strength, and it is not yet possible to fully predict this micelle-like assembly [38]. However, unlike most globular proteins, capsid proteins self-assemble by design. The experimental reproducibility and robustness is thus high, making their formation less problematic to analyze. Given the relatively small number of molecules involved, viral capsid assembly is also more amenable to simulations with molecular-scale details than the crystallization of globular proteins, for example. As a result, there is significantly more connection between experiments (which stimulate computational work) and simulations (which lead to experimentally testable predictions) than for other protein self-assembly processes [2••].

Minimal models of capsid proteins, like those described earlier for globular proteins, require an interaction anisotropy and specificity to capture qualitative features of their assembly [39]. With this minimal set of features, simulations and experiments show that the assembly of empty capsids proceeds via nucleation and growth, which is analogous to the crystallization of globular proteins [39]. For both processes, the loss of translational entropy during self-assembly is offset by gains from specific hydrophobic, electrostatic, van der Waals, and hydrogen-bonding interactions. The non-planar geometry of the capsid does influence the self-assembly process (and may alter the conformation of the proteins involved). Analogous experiments at larger length scales show that crystallization of colloidal particles is also altered on the surface of a sphere [40].

In the presence of a polynucleotide, electrostatic interactions between positive charges on the capsid protein and negative charges on the nucleic acids can promote or even dominate the capsid assembly process [2••]. The length of the polynucleotide encapsulated is also important, since longer RNA or DNA strands are encapsulated at high entropic cost while shorter strands may not provide sufficient electrostatic stabilization [41••]. The optimal nucleic acid length for some viral capsids has been measured [35]. The self-assembly properties of this model are in excellent agreement with experimental results for a variety of viruses.

An ongoing challenge related to viral self-assembly is the characterization of the interaction anisotropy that results in both protein–protein and protein–genome interactions. A further challenge is to identify and characterize intermediate structures along the assembly pathways. It is interesting to note that a similar challenge is faced in the study of protein crystallization, where two main ordering pathways have been proposed: direct nucleation from a solution of monomers, and the formation of long-lived metastable protein clusters from which a crystal emerges [42••].

4. Amyloid fibrils

Amyloid fibrils are insoluble protein aggregates with a cross-β structure in which β-strands form almost continuous hydrogen-bonded β-sheets that run along the fibril. The fibrils formed from different proteins are qualitatively similar: unbranched filamentous structures that are a few nanometers in diameter but can grow to be several microns in length. This common structure is likely driven by the universal tendency of polypeptide chains to form hydrogen bonds between atoms along the backbone [43]. For peptides, other hierarchical twisted assemblies, such as ribbons, fibrils, and stacks (or fibers) [44], are possible, and several types of peptide nanotubes have been described [45]. It has been suggested that the fibril can be thought of as a one-dimensional pseudo-crystal, and in that sense, the fibril is the most organized structure that a flexible polypeptide chain can form. It has also been speculated that under certain conditions, an amyloid fibril made of short polypeptides (<150 residues) would be thermodynamically more stable than even the functional native state [43]. This proposal could help explain why most amyloid-related diseases are caused by

short peptides or proteins. It also implies that understanding the kinetics of amyloid formation is essential for disease prevention.

The most-studied protein that forms amyloid fibril is amyloid β (Aβ), a peptide of 40–42 amino acids associated with the pathogenesis of Alzheimer's disease [46]. A combination of experiments and simulation indicate that, just as for protein crystallization and viral capsid assembly, there are two main pathways for the condensed phase to form: the fibrils may (i) directly nucleate from a solution of Aβ monomers or (ii) first self-assemble into an oligomer from which the fibril forms in a second nucleation event [7]. In both cases, there is a conformation change that occurs as the fibril forms. In addition, existing fibrils can seed the formation of new fibrils that then subsequently break off from the original fibril [7].

When fully atomistic models are used, it is possible to obtain reasonable agreement between simulations and experiments of Aβ fibrilization [47]. Coarse-grained models based on patchy particles have also been proposed [48–51]. However, the factors that control the specific pathway that is observed under a given set of conditions remain poorly understood. The fibrilization is strongly dependent on ionic strength and on the identity of the peptide, and it is currently not possible to predict the pathway that will be taken by a specific peptide under a given set of conditions [52].

Despite considerable effort, a number of issues remain unresolved. It is increasingly accepted that for Aβ, oligomeric peptide assemblies are more neurotoxic than fibril plaques, yet the links between the kinetics and mechanisms for fibril growth in vitro and those in vivo are not always clear [46]. For Aβ and other amyloid-forming peptides and proteins, cell membranes may be important in the nucleation process [53,54]. Since much of the work on amyloid-forming peptides and proteins is driven by a search for therapies to prevent, slow down or even stop the course of amyloid-related diseases, there is a pressing need to develop a complete picture of the self-assembly of amyloid-forming proteins and peptides.

5. Conclusions and future directions

Protein self-assembly plays an important role in numerous biophysical process. Here we selected three systems—globular proteins, viral capsids, and amyloid fibrils—to illustrate recent progress and current challenges in understanding protein self-assembly from a soft matter viewpoint. We conclude by highlighting several topics that in our opinion should be further investigated.

5.1. Solvent effects

A comprehensive understanding of the role of the solvent in protein self-assembly has yet to be achieved. In aqueous solution, the role of pH and ionic strength has been considered for many proteins and explained using both isotropic and anisotropic colloidal models [55–57]. The addition of other solvents (and small molecules) such as glycerol, carbohydrates, amino acids, and nucleic acids can have a profound influence on protein self assembly (either by direct interaction with the protein surface, or by modifying the solvent characteristics) [12••,43,58]. These effects are particularly important when formulating proteins for liquid storage or lyophilization and further understanding of the mechanisms for self-assembly would improve the success of these processes [11]. Crystallization, for instance, can sometimes be improved by the addition of glycerol, which is thought to suppress nucleation and thus results in the formation of fewer, larger crystals [59].

Simulations are a natural way to analyze solvent effects as it is possible to treat the solvent explicitly. Given the computational cost of working with explicit solvent, a multiscale approach is usually needed: two proteins are simulated in a solvent to deduce the parameters of an effective pair interaction energy that is then used to study self-assembly in a solvent-free system [49•]. Another approach uses a coarse-grained optimized potential for efficient protein structure prediction (OPEP)

combined with hydrodynamics [60] and may be a useful computational tool in the future. There are cases, however, for which it is not possible to separate out the solvent without sacrificing essential details, such as for membrane proteins, which can only exist in their native state when embedded in a lipid membrane or solubilized in a detergent. Unsurprisingly, our understanding of the self-assembly of membrane proteins is still in its infancy.

5.2. Small molecules

The conjugation of small molecules to proteins is routinely performed to conduct analytical testing (e.g., fluorescent tagging), to improve the biological compatibility, (e.g., PEGylation) or to develop a new therapeutic product, (e.g. in a protein-drug conjugate) [61•–63]. In each case, a small molecule is covalently attached to a protein to modify its behavior. This in turn can alter the self-assembly [61•]. Conjugation typically occurs at either a primary amine or at a free cysteine, although specific chemistry to modify proteins at other amino acids has also been developed [64•]. While these strategies are often used, there are far fewer experimental, simulation, or theoretical descriptions of how these modifications alter the protein self-assembly than for unmodified proteins.

5.3. Non-compact proteins

When modeling proteins for coarse-grained simulations of self-assembly, it is usually assumed that shape anisotropy is negligible, with exceptions only when the anisotropy is too pronounced (e.g., spherocylinders to model Aβ [49•]). Although this may be a reasonable assumption for most single-domain proteins, for multi-domain proteins it must be used with care, and there are families of proteins, for which the isotropic-shape approximation is a gross oversimplification.

5.4. Kinetics

Much of the work on protein self-assembly has focused on the thermodynamic behavior, but kinetics often also plays a crucial role in determining the outcome of assembly [65•–67]. The most notorious example is that of protein crystallization. Although the crystal may be the most stable state under the conditions studied (as can be verified by seeding the solution with a crystal and watching it grow), a crystal may not form spontaneously, even after many months, if the nucleation rate is slow [68•,69•]. In order to obtain a comprehensive understanding of the phenomenon (in particular, of what will happen for a given set of conditions), additional computational work should be carried out to examine the kinetics of self-assembly for particles with anisotropic interactions and connect the results with experimental data on nucleation and growth of self-assembled phases.

5.5. Purposeful vs. incidental self-assembly

As we mentioned in the introduction, some proteins self-assemble by design while others do so only when things go wrong or the solution conditions are perturbed. In other words, the structure of some proteins is such that self-assembly occurs for a specific purpose (such as to encapsulate other molecules), while for other proteins, self-assembly is not central to their function—which they carry out in the unassembled state—and only occurs because of incidental physical considerations. It would be interesting to compare the properties of proteins involved in purposeful and incidental processes to see whether any new insights may be gleaned regarding the specific and universal features of protein self-assembly (Fig. 3).

We hope that this brief overview of the physics of protein self-assembly will stimulate others to tackle some of the outstanding issues in the field.
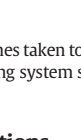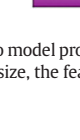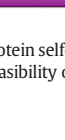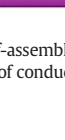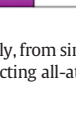
MODEL DESCRIPTION	MODEL ILLUSTRATION	TYPES OF TACKLED PROTEIN ASSEMBLIES			
		GLOBULAR	AMYLOIDS	VIRUSES	MEMBRANE
ISOTROPIC ATTRACTION		YES	YES	YES	YES
PATCHY SPHERE		YES	YES	YES	NO
ONE-TO-ONE PATCHY		YES	YES	NO	NO
PATCHY ROD/ELLIPSOID		YES	YES	NO	NO
ATOMISTIC WITH IMPLICIT SOLVENT		YES	YES	YES	YES
ATOMISTIC WITH EXPLICIT SOLVENT		NO	NO	NO	YES

Fig. 3. The approaches taken to model protein self-assembly, from simple isotropic models to all atom simulations with explicit solvent, have been used to support or explain experimental data. With increasing system size, the feasibility of conducting all-atom simulations with explicit solvent reduces significantly.

6. Future directions

- A more comprehensive view of the kinetics and intermediate pathways of protein self-assembly.
- Minimal models for protein self-assembly that do not require crystal structures as experimental inputs.
- A better understanding of how molecular anisotropy directs protein assembly.
- Greater use of atomistic models with explicit solvent to describe protein-protein interactions.

References recommended reading[*][**]

[1] Ahnert SE, Marsh JA, Hernández H, Robinson CV, Teichmann SA. Principles of assembly reveal a periodic table of protein complexes. Science 2015;350:6266. http://dx.doi.org/10.1126/science.aaa2245.

[2••] Hagan MF. Modeling viral capsid assembly. Adv Chem Phys 2014;155:1–68. http://dx.doi.org/10.1002/9781118755815.ch01.[Review of the recent advances in the physics of virus capsid assembly].

[3••] Yeates TO, Crowley CS, Tanaka S. Bacterial microcompartment organelles: protein shell structure and evolution. Ann Rev Biophys 2010;39:185–205. http://dx.doi.org/10.1146/annurev.biophys.093008.131418.[A detailed presentation of the protein-based structures found in bacteria that serve as organelles for specific metabolic pathways]

[4•] Knowles TPJ, Vendruscolo M, Dobson CM. The amyloid state and its association with protein misfolding diseases. Nat Rev Mol Cell Biol 2014;15:384–96. http://dx.doi.org/10.1038/nrm3810 [This review article summarizes the current understanding of the thermodynamics and kinetics of amyloid formation. It discusses the general features of amyloid structures and offers a basis for therapeutic advances in treating amyloid diseases].

[5••] Eaton WA, Hofrichter J. Sickle cell hemoglobin polymerization. Adv Protein Chem 1990;40:63–279.[A classic, comprehensive review of sickle cell hemoglobin self-assembly].

[6••] Luft JR, Wolfley JR, Snell EH. What's in a drop? Correlating observations and outcomes to guide macromolecular crystallization experiments. Cryst Growth Des 2011;11(3):651–63. http://dx.doi.org/10.1021/cg1013945.[An exhaustive summary of the possible outcomes of protein crystallization experiments. The article offers a comprehensive series of images illustrating the many results of protein self-assembly]

[7] Cohen SIA, Linse S, Luheshi LM, Hellstrand E, White DA, Rajah L, Otzen DE, Vendruscolo M, Dobson CM, Knowles TPJ. Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism. Proc Natl Acad Sci U S A 2013;110(24):9758–63. http://dx.doi.org/10.1073/pnas.1218402110.

[8] Benedek GB. Cataract as a protein condensation disease: the Proctor Lecture. Invest Ophthalmol Vis Sci 1997;38(10):1911–21.

[9] Uversky VN, Eliezer D. Biophysics of Parkinson's disease: structure and aggregation of α-synuclein. Curr Protein Pept Sci 2009;10(5):483–99. http://dx.doi.org/10.2174/138920309789351921.

[10] Flickinger MC, editor. Downstream Industrial Biotechnology: recovery and Purification. Hoboken: Wiley; 2013.

[11] Carpenter JF, Manning MC, editors. Rational design of stable protein formulations: theory and practice. New York: Springer; 2002.

[12••] McPherson A, Gavira JA. Introduction to protein crystallization. Acta Crystallogr F Struct Biol Commun 2014;70(1):2–20. http://dx.doi.org/10.1107/S2053230X13033141. [A thorough overview of all aspects of protein crystallization starting with the history of the field over 150 years ago to the latest advances and future directions]

[13•] Dumetz AC, Chockla AM, Kaler EW, Lenhoff AM. Protein phase behavior in aqueous solutions: crystallization, liquid–liquid phase separation, gels, and aggregates. Biophys J 2008;94(2):570–83. http://dx.doi.org/10.1529/biophysj.107.116152. [This article reports the various self-assembled states that are found in six proteins (ovalbumin, ribonuclease A, soybean trypsin inhibitor, lysozyme, and β-lactoglobulin A and B) at different temperatures and salt conditions. It provides phase boundaries and photographs of the different phases that form].

[14] Asherie N, Ginsberg C, Blass S, Greenbaum A, Knafo S. Solubility of thaumatin. Cryst Growth Des 2008;8(6):1815–7. http://dx.doi.org/10.1021/cg800276r.

[15] Veesler S, Fertè N, Costes M-S, Czjzek M, Astier J-P. Temperature and pH Effect on the polymorphism of aprotinin (BPTI) in sodium bromide solutions. Crystal Growth Des 2004;4(6):1137–41. http://dx.doi.org/10.1021/cg0498195.

---

[*] of special interest.
[**] of outstanding interest.

[16] Asherie N, Jakoncic J, Ginsberg C, Greenbaum A, Stojanoff V, Hrnjez BJ, Blass S, Berger J. Tartrate chirality determines thaumatin crystal habit. Cryst Growth Des 2009;9(9): 4189–98. http://dx.doi.org/10.1021/cg900465h.

[17] James S, Quinn MK, McManus JJ. The self assembly of proteins; probing patchy protein interactions. Phys Chem Chem Phys 2015;17:5413–20. http://dx.doi.org/10.1039/C4CP05892E.

[18] Galkin O, Vekilov PG. Control of protein crystal nucleation around the metastable liquid–liquid phase boundary. Proc Natl Acad Sci U S A 2000;97:6277–81.

[19] Berland CR, Thurston GM, Kondo M, Broide ML, Pande J, Ogun O, Benedek GB. Solid–liquid phase boundaries of lens protein solutions. Proc Natl Acad Sci U S A 1992; 89(4):1214–8.

[20] Grouzael SF, Bonneté Astier JP, Ferté N, Perez J, Veesler S. Exploring bovine pancreatic trypsin inhibitor phase transitions. J Phys Chem B 2006;110:19664–70. http://dx.doi.org/10.1021/Jp0627123.

[21] Anderson VJ, Lekkerkerker HNW. Insights into phase transition kinetics from colloid science. Nature 2002;416:811–5.

[22] Foffi G, McCullagh GD, Lawlor A, Zaccarelli E, Dawson KA, Sciortino F, Tartaglia P, Pini D, Stell G. Phase equilibria and glass transition in colloidal systems with short-ranged attractive interactions. Applications to protein crystallization. Phys Rev E 2002;65:031407.

[23] Cardinaux F, Gibaud T, Stradner A, Schurtenberger P. Interplay between spinodal decomposition and glass formation in proteins exhibiting short-range attractions. Phys Rev Lett 2007;99(11):118301. http://dx.doi.org/10.1103/PhysRevLett.99.118301.

[24•] Piazza R. Protein interactions and association: an open challenge for colloid science. Curr Opin Coll Int Sci 2004;8(6):515–22. http://dx.doi.org/10.1016/j.cocis.2004.01.008. [A review about the use of simple colloidal models to explain protein phase behavior. A useful starting point to see the development of ideas over the last decade].

[25••] Lomakin A, Asherie N, Benedek GB. Aeolotopic interactions of globular proteins. Proc Natl Acad Sci 1999;96(17):9465–8. http://dx.doi.org/10.1073/pnas.96.17.9465. [One of the first works carefully documenting the impact of interaction anisotropy on the protein phase behavior].

[26•] Sear RP. Phase behavior of a simple model of globular proteins. J Chem Phys 1999; 111(10):4800–6.[Another early contribution on the role of interaction anisotropy on protein phase behavior].

[27•] Fusco D, Charbonneau P. Soft matter perspective on protein crystal assembly. Coll Surf B: Biointerfaces 2016;137:22–31. http://dx.doi.org/10.1016/j.colsurfb.2015.07.023. [Recent review of interaction anisotropy and its impact on protein crystal assembly]

[28] Finkelstein AV, Ptitsyn O. Protein physics: a course of lectures. London: Academic Press; 2002.

[29•] Kern N, Frenkel D. Fluid–fluid coexistence in colloidal systems with short-ranged strongly directional attraction". J Chem Phys 2003;118:9882. http://dx.doi.org/10.1063/1.1569473. [Introduced the most widely used patchy particle model employed in simulations allowing bond distances and bond angles to be tuned separately]

[30•] Bianchi E, Largo J, Tartaglia P, Zaccarelli E, Sciortino F. Phase diagram of patchy colloids: towards empty liquids. Phys Rev Lett 2006;97:168301. http://dx.doi.org/10.1103/PhysRevLett.97.168301. [The first report of the systematic shift of the phase coexistence at lower densities with valence in the one-bond-per-patch limit which is relevant for proteins]

[31] Liu H, Kumar SK, Sciortino F. Vapor–liquid coexistence of patchy models: Relevance to protein phase behavior. J Chem Phys 2007;127:084902. http://dx.doi.org/10.1063/1.2768056.

[32] Fusco D, Barnum TJ, Bruno AE, Luft JR, Snell EH, Mukherjee S, Charbonneau P. Statistical analysis of crystallization database links protein physico-chemical features with crystallization mechanisms. PLoS One 2014;9(7):101123. http://dx.doi.org/10.1371/journal.pone.0101123.

[33] Altan I, Charbonneau P, Snell EH. Computational crystallization. Arch Biochem Biophys 2016. http://dx.doi.org/10.1016/j.abb.2016.01.004 (in press).

[34•] Goldschmidt L, Cooper DR, Derewenda ZS, Eisenberg D. Toward rational protein crystallization: AaWeb server for the design of crystallizable protein variants. Protein Sci 2007;16(8):1569–76. http://dx.doi.org/10.1110/ps.072914007. [An in-depth explanation of the algorithm used to offer recommendations for residues to be mutated in order to increase the likelihood of crystallizing a protein of known sequence. The algorithm builds upon the surface entropy reduction approach pioneered by Z.S. Derewenda.].

[35] Cadena-Nava RD, Comas-Garcia M, Garmann RF, Rao ALN, Knobler CM, Gelbart WM. Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. J Virol 2012;86(6):3318–26. http://dx.doi.org/10.1128/JVI.06566-11.

[36] Castón JR, Carrascosa JL The basic architecture of viruses in structure and physics of viruses: an integrated textbook, volume 68 of the series Subcellular Biochemistry. In: Mateu MG, editor. Dordrecht: Springer; 2013.

[37] Lavelle L, Michel J-P, Gingery M. The disassembly, reassembly and stability of CCMV protein capsids. J Virol Meth 2007;146:311–6. http://dx.doi.org/10.1016/j.jviromet.2007.07.020.

[38] Comas-Garcia M, Garmann RF, Singaram SW, Ben-Shaul A, Knobler CM, Gelbart WM. Characterization of viral capsid protein self-assembly around short single-stranded rna. J Phys Chem B 2014;118(27):7510–9. http://dx.doi.org/10.1021/jp503050z.

[39] Perlmutter JD, Hagan MF. Mechanisms of Virus Assembly. Annu Rev Phys Chem 2015;66:217–39. http://dx.doi.org/10.1146/annurev-physchem-040214-121637.

[40] Meng G, Paulose J, Nelson DR, Manoharan VN. Elastic instability of a crystal growing on a curved surface. Science 2014;343(6171):634–7. http://dx.doi.org/10.1126/science.1244827.

[41••] Garmann RF, Comas-Garcia M, Knobler CM, Gelbart WM. Physical principles in the self-assembly of a simple spherical Virus. Acc Chem Res 2016;49(1):48–55.

http://dx.doi.org/10.1021/acs.accounts.5b00350. [This article discusses three key paramaters—pH, RNA length and RNA-protein charge ratio—in the context of a physical chemistry approach to the self-assembly of cowpea chlorotic mottle virus]

[42••] Vekilov PG. Nucleation. Cryst Growth Des 2010;10(12):5007–19. http://dx.doi.org/10.1021/cg1011633. [A review of nucleation of crystals from solution with an emphasis on proteins]

[43] Buell AK, Galvagnion C, Gaspar R, Sparr E, Vendruscolo M, Knowles TPJ, Linse S, Dobson CM. Solution conditions determine the relative importance of nucleation and growth processes in α-synuclein aggregation. Proc Natl Acad Sci U S A 2014; 111:7671–6. http://dx.doi.org/10.1073/pnas.1315346111.

[44] Aggeli A, Nyrkova IA, Bell M, Harding R, Carrick L, McLeish TCB, Semenov AN, Boden N. Hierarchical self-assembly of chiral rod-like molecules as a model for peptide β-sheet tapes, ribbons, fibrils, and fibers. Proc Natl Acad Sci U S A 2001;98(21): 11857–62. http://dx.doi.org/10.1073/pnas.191250198.

[45] Valéry C, Artzner F, Paternostre M. Peptide nanotubes: Molecular organisations, self-assembly mechanisms and applications. Soft Matter 2011;7(20):9583–94. http://dx.doi.org/10.1039/c1sm05698k.

[46] Roychaudhuri R, Yang M, Minako MH, Teplow DB. Amyloid β-protein assembly and Alzheimer disease. J Biol Chem 2009;284:4749–53. http://dx.doi.org/10.1074/jbc.R800036200.

[47] Viet MH, Nguyen PH, Derreumaux P, Li MS. Effect of the english familial disease mutation (H6R) on the monomers and dimers of Aβ40 and Aβ42. ACS Chem Neurosci 2014;5(8):646–57. http://dx.doi.org/10.1021/cn500007j.

[48] Bieler NS, Knowles TPJ, Frenkel D, Vácha R. Connecting macroscopic observables and microscopic assembly events in amyloid formation using coarse grained simulations. PLoS Comput Biol 2012;8(10), e1002692. http://dx.doi.org/10.1371/journal.pcbi.1002692.

[49•] Šarić A, Chebaro YC, Knowles TPJ, Frenkel D. Crucial role of nonspecific interactions in amyloid nucleation. Proc Natl Acad Sci U S A 2014;111:17869. http://dx.doi.org/10.1073/pnas.1410159111. [Recent advances in modelling amyloid assembly]

[50] Ilie IM, den Otter WK, Briels WJ. Rotational Brownian dynamics simulations of clathrin cage formation. J Chem Phys 2014;141:065101. http://dx.doi.org/10.1063/1.4891306.

[51] Vissers T, Smallenburg F, Gianmarco Munao, Preisler Z, Sciortino F. Cooperative polymerization of one-patch colloids. J Chem Phys 2014;140:144902. http://dx.doi.org/10.1063/1.4869834.

[52] Meisl G, Yang X, Frohm B, Knowles TPJ, Linse S. Quantitative analysis of intrinsic and extrinsic factors in the aggregation mechanism of Alzheimer-associated Aβ-peptide. Sci Rep 2016;6:18728. http://dx.doi.org/10.1038/srep18728.

[53] Tofoleanu F, Viorel Buchete N. Alzheimer Aβ peptide interactions with lipid membranes: Fibrils, oligomers and polymorphic amyloid channels. Prion 2012; 6(4):339–45. http://dx.doi.org/10.4161/pri.21022.

[54] van Rooijen BD, Claessens MMAE, Subramaniam V. Membrane Interactions of Oligomeric Alpha-Synuclein: Potential Role in Parkinson's Disease. Curr Prot Pep Sci 2010;11(5):334–42. http://dx.doi.org/10.1371/journal.pone.0014292.

[55] Roberts D, Keeling R, Tracka M, Wan Der Walle CF, Uddin S, Warwicker J, Curtis R. Specific ion and buffer effects on protein-protein interactions of a monoclonal antibody. Mol Pharm 2015;12(1):179–93. http://dx.doi.org/10.1021/mp500533c.

[56] Li W, Persson BA, Morin M, Behrens MA, Lund M, Zackrisson Oskolkova M. Charge-induced patchy attractions between proteins. J Phys Chem B 2015;119(2):503–8. http://dx.doi.org/10.1021/jp512027j.

[57] Kastelic M, Kalyuzhnyi YV, Hribar-Lee B, Dill KA, Vlachy V. Protein aggregation in salt solutions. Proc Natl Acad Sci U S A 2015;112(21):6766–70. http://dx.doi.org/10.1073/pnas.1507303112.

[58] Blumlein A, McManus JJ. Reversible and non-reversible thermal denaturation of lysozyme with varying pH at low ionic strength. Biochim Biophys Protein Proteomics 2013;1834(10):2064–70. http://dx.doi.org/10.1016/j.bbapap.2013.06.001.

[59] Benvenuti M, Mangani S. Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. Nat Protoc 2007;2:1633–51. http://dx.doi.org/10.1038/nprot.2007.198.

[60] Sterpone F, Derreumaux P, Melchionna S. Protein simulations in fluids: coupling the OPEP coarse-grained force field with hydrodynamics. J Chem Theory Comput 2015; 11:1843–53. http://dx.doi.org/10.1021/ct501015h.

[61•] Quinn MK, Gnan N, James S, Ninarello A, Sciortino F, Zaccarelli E, McManus JJ. How fluorescent labelling alters the solution behaviour of proteins. Phys Chem Chem Phys 2015;17:31177–87. http://dx.doi.org/10.1039/c5cp04463d. [A recent report on how chemical modification can alter the self-assembly of proteins, using a combination of experiments and simulations to understand protein phase behavior]

[62] Cattani G, Vogeley L, Crowley PB. Structure of a PEGylated protein reveals a highly porous double-helical assembly. Nat Chem 2015;7:823–8. http://dx.doi.org/10.1038/nchem.2342.

[63] Adem YT, Schwarz KA, Duenas E, Patapoff TW, Galush WJ, Esue O. Auristatin antibody drug conjugate physical instability and the role of drug payload. Bioconjug Chem 2014;25(4):656–64. http://dx.doi.org/10.1021/bc400439x.

[64•] Boutureira O, Bernardes GJL. Advances in chemical protein modification. Chem Rev 2015;115:2174–95. http://dx.doi.org/10.1021/cr500399p.[A recent review of chemical approaches for performing site-selective chemical modifications to proteins using metal mediated and metal-free methods]

[65•] Whitelam S. Control of pathways and yields of protein crystallization through the interplay of nonspecific and specific attractions. Phys Rev Lett 2010;105(8):088102. http://dx.doi.org/10.1103/PhysRevLett.105.088102. [Numerical work which identifies different dynamic pathways to protein crystallization by tuning the strength of effective anisotropic and isotropic interactions].

[66] Haxton TK, Whitelam S. Design rules for the self-assembly of a protein crystal. Soft Matter 2012;8(13):3558-356. http://dx.doi.org/10.1039/c2sm07436b.

[67] Newton AC, Groenewold J, Kegel WK, Bolhuis PG. Rotational diffusion affects the dynamical self-assembly pathways of patchy particles. Proc Natl Acad Sci U S A 2015; 112(50):15308–13. http://dx.doi.org/10.1073/pnas.1513210112.

[68•] Budayova-Spano M, Dauvergne F, Audiffren M, Bactiveland T, Cusack S. A methodology and an instrument for the temperature-controlled optimization of crystal growth. Acta Cryst 2007;D63:339–47. http://dx.doi.org/10.1107/S0907444906054230. [Proposes an efficient experimental protocol for protein crystallization].

[69•] Heymann M, Opthalage A, Wierman JL, Akella S, Szebenyi DME, Gruner SM, Fraden S. Room-temperature serial crystallography using a kinetically optimized microfluidic device for protein crystallization and on-chip X-ray diffraction. IUCrJ 2014;1(5): 349–60. http://dx.doi.org/10.1107/S2052252514016960. [Describes an experimental method to perform serial crystallography on crystals grown from nano-litre sized droplets].