Soft Matter

PAPER

View Article Online
View Journal | View Issue

Cite this: *Soft Matter*, 2015, **11**, 2036

Implications of protein polymorphism on protein phase behaviour†

J. Stegen*[ab] and P. van der Schoot[ab]

The phase behaviour of small globular proteins is often modeled by approximating them as spherical particles with fixed internal structure. However, changes in the local environment of a protein can lead to changes in its conformation rendering this approximation invalid. We present a simple two-state model in which protein conformation is not conserved and where the high-energy, non-native state is stabilised by pair-wise attractive interactions. The resulting phase behaviour is remarkably complex, non-universal and exhibits re-entrance. The model calculations show a demarcation between a regime where conformational transitioning is largely enslaved by phase separation and one where this is not the case. In the latter regime, which is characterised by a large free energy difference between the native and the non-native state, we deduce that the kinetics of the phase transition strongly depend on the average conformation of the proteins prior to their condensation. For condensation to occur in this regime within a dispersion of native proteins, nucleation of a cluster of proteins in the non-native state is required. We argue that our theory supports the distinction between common phase separation and the nucleated assembly of non-native supramolecular aggregates in protein dispersions.

Received 1st January 2015
Accepted 19th January 2015

DOI: 10.1039/c5sm00003c

www.rsc.org/softmatter

1 Introduction

A large body of work, theoretical and experimental, has been devoted to studying the phase behaviour of globular proteins, reflecting the complexity and relevance of the subject. An understanding of protein phase behaviour is relevant in the context of the food[1–4] and pharmaceutical industry,[5,6] for the structural characterisation of proteins,[7] for understanding the behaviour of proteins within the crowded environment of a cell,[8,9] as well as for understanding numerous neurodegenerative diseases that have been linked to the formation of amyloid fibrils.[10] Furthermore, an understanding of protein dispersions is relevant to the design of bio-based and biomimetic molecular materials,[11] and due to its complexity it is of inherent scientific interest.

Proteins[12] have a heterogeneous surface that is different for every kind of protein, which gives rise to inherently anisotropic interactions with other proteins, even if they are globular.[13–16] The dependence of the surface properties and the charge distribution on the pH, and the dependence of electrostatic interactions on the ionic strength of the solution[17–20] add to the complexity.[5,21] In fact, proteins are never perfectly spherical, are in principle deformable[22] and in many cases are able to form supramolecular structures in native and non-native conformational states.[23–26] All of this translates into very rich phase behavior,[27] including liquid–liquid phase separation, crystallization, gelation and aggregation into a variety of different supramolecular structures.[23–26]

The coupling of conformation to phase behaviour of peptides has received some attention,[28,29] however the consequences of possible changes in protein conformation on liquid–liquid phase separation has to our knowledge received no prior attention. Inspired by a two-state model for polymer phase behaviour,[30] we theoretically address this topic. We do so by presuming a competition between two conformational states, being the native and a *single* non-native state that we do not specify. This differs from current models, which either presume the protein to remain in the native state and macroscopic phase behaviour to be reversible,[31] or presume the native state to be non-conserved with corresponding irreversible phase and/or aggregation behaviour.[26] We do not make this distinction here, but rather presume that changes in protein conformation in principle occur on a continuous scale ranging from being negligible to the complete loss of the native structure. We postulate that selection of the dominant non-native conformer is driven by the strength of the interactions between them and that these interactions ultimately drive condensation of the proteins. Whilst we presume protein phase behaviour to be reversible, we deduce from our model that reversibility is in some cases exceedingly slow and should not be observed on experimental time scales.

[a]*Applied Physics, Eindhoven University of Technology, Eindhoven, 5600 MB, The Netherlands. E-mail: j.stegen@tue.nl*

[b]*Institute for Theoretical Physics, Utrecht University, Leuvenlaan 4, 3584 CE, Utrecht, The Netherlands*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5sm00003c

We presume that the change in protein conformation is the result of a change in the local microscopic environment of the protein, that is, it does not result from the hydrolysis of the protein into short polypeptides which is irreversible. Indeed, it is well known that the local environment can affect protein structure and that proteins must have some 'conformational softness' to perform their function within a living organism.[32–34] For instance, the adsorption of a protein at an interface almost invariantly result in conformational changes in the protein.[35,36] Also, changes in solution conditions can affect the protein conformation and even protein concentration alone affects the conformation of some proteins.[37] Here, we consider a system similar to this and let protein concentration be the source of changes in local environment. Through mass action, an increasing number of protein–protein interactions with increasing concentration, lead to an increasing number of proteins in the non-native state.

In principle, at fixed concentration such changes in local environment can result from a phase transition or local density fluctuations. Here, we run into a dilemma, which must come first? Does a phase transition induce changes in protein structure or do changes in protein structure induce a phase transition? Our work shows that the answer depends on the degree to which the conformation of the protein changes. For small changes in protein conformation, that we translate into a low free energy penalty relative to the native state, the phase transition induces a change in protein conformation. However for large changes in structure, *i.e.*, a large free energy penalty, the new non-native conformation must typically be nucleated before a macroscopic phase transition (condensation) can occur. The two thermodynamic models that we invoke produce consistent predictions, suggesting that our conclusions are robust.

The remainder of this paper is structured as follows. In Section 2 we introduce our two-state protein model in which high free energy conformers are stabilised by attractive interactions with other high free energy conformers, and discuss its physical interpretation. In Section 3 we show that a first-order conformational phase transition between a dispersion state with proteins mostly in their native state and a dispersion state with most proteins in their non-native state exists. In Section 4 we combine our two-state protein model with a Flory–Huggins-type free energy[38,39] and with a Carnahan–Starling-based free energy[40] and discuss how phase and stability diagrams can be constructed.

In Sections 5 and 6 we present and discuss the corresponding phase and stability diagrams where we show that the phase behaviour exhibits re-entrance and is non-universal, *i.e.*, a law of corresponding states does not exist. Furthermore, we discuss the thermodynamic stability of the dispersion with respect to phase separation where, based on the local curvature of the free energy surface, we distinguish between spinodal decomposition and nucleation and growth. Additionally, we consider conformational relaxation of the proteins and show in Section 6 that the presence of a first-order conformational phase transition causes the thermodynamic stability to become inherently dependent on the protein conformation.

Finally, in Section 7, we provide an in-depth analysis of the predictions of our models to show that even if one presumes that changes in protein conformation occur on a continuous energetic scale, one finds a clear demarcation between a regime where the coupling between the phase behaviour and changes in protein conformation is weak, *i.e.*, changes in protein structure are induced by a phase transition and have a negligible impact on phase behaviour, and a regime where the coupling is strong and the opposite holds true.

2 Two-state protein model

Our two-state protein model is based on the presumption that a protein can either be in its native state or in its non-native state, and that the protein can reversibly switch between these two states. There is a free energy difference ε (in units of thermal energy, k_BT) between these two states. There are steric (excluded-volume) interactions between all proteins, and proteins that are in the non-native state can engage in an attractive interaction with nearby proteins that are also in the non-native state. The strength of this interaction is measured by a dimensionless, so-called Flory parameter χ. All proteins that are in the non-native state are in an *identical*, non-native state. The model is graphically summarised in Fig. 1.

The thermodynamic state of the protein dispersion depends, apart from the two energetic parameters, on the total volume fraction of protein, ϕ, which equals the protein number density times the volume of a protein, v_0, and the number fraction η of the proteins in the non-native state. Tacit assumption is that the volume of a protein is equal in both states, while this must not necessarily be the case, we presume this to be true for reasons of simplicity. In future work we will address the situation where the native and non-native conformation are not of equal effective volume. So, if η is the number fraction of proteins in the non-native state, then $\eta\phi$ is the volume fraction of the solution occupied by proteins in the non-native state, while $(1-\eta)\phi$ is the volume fraction occupied by proteins in the native state. The volume fraction of solvent equals $1-\phi$.

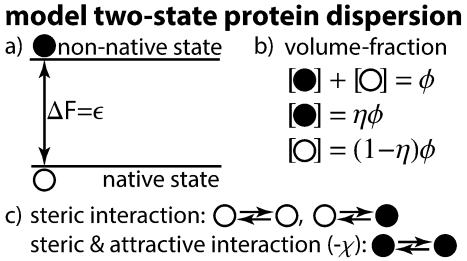

Fig. 1 Ingredients of the two-state model. (a) The native state (open circle) and non-native state (filled circle) of the model two-state protein are separated by a free energy difference ε (in units thermal energy). (b) The overall protein volume fraction is ϕ, the fraction of proteins in the non-native state is η. (c) The nature of protein–protein interactions depends on the conformational state of the proteins. All proteins interact *via* excluded volumes and nearby pairs of protein that are in the non-native state engage in attractive interactions of strength $-\chi$ (in units thermal energy).

We are now in a position to construct our free energy. Let N denote the total number of proteins in the dispersion. The total number of proteins in the non-native state must then be equal to $N\eta$, while the total number of proteins in the native state is $N(1-\eta)$. In the mean-field approximation the distribution of native and non-native states over the N proteins is independent, allowing us to directly write down the Gibbsian entropy due to the increased number of microscopic states available to the dispersion. The corresponding entropic contribution to the dimensionless free energy density is,

$$f_{\text{entr}} = \eta\phi \ln \eta + (1-\eta)\phi \ln 1-\eta, \tag{1}$$

where f_{entr} is scaled to the volume of a single protein, v_0, and in units of thermal energy, k_BT. The enthalpic contribution to the free energy, we calculate next.

Having $N\eta$ proteins in the non-native state comes at a free energy cost of $N\eta\varepsilon$, which corresponds to an increase of the dimensionless free energy density of $\eta\phi\varepsilon$. The free energy cost of having proteins in the non-native state can be compensated for by attractive interactions between them of strength $-\chi$. From mean-field arguments it follows that the dimensionless free energy density associated with these interactions is $-\chi(\eta\phi)^2/2$, because the probability that two proteins that are in the non-native state and in range of each others attractive potential is proportional to $(\eta\phi)^2$, while the factor of 1/2 corrects for double counting. Combining the two energetic contributions we obtain the following dimensionless free energy density,

$$f_{\text{enth}} = \varepsilon\eta\phi - \frac{\chi}{2}(\eta\phi)^2. \tag{2}$$

The total contribution to the dimensionless free energy density of our two-state protein model is now given by,

$$f_{2s} = f_{\text{entr}} + f_{\text{enth}}. \tag{3}$$

In summary, the state of the protein dispersion is characterised by two order parameters, ϕ and η, while the external conditions and the protein properties determine the values of the two energetic parameters, χ and ε.‡ Because we lack a microscopic model for the latter two, we cannot predict how these respond to changes in say temperature and other solution conditions including ionic strength and acidity. Hence, we treat χ and ε as phenomenological parameters, also to keep the theory as general as possible. Although the phase behaviour predicted by our model is a function of χ and ε alone and not of the underlying microscopic model relating these parameters to solution conditions, these parameters are in principle amenable to experimental determination as a function of solution conditions.[17,20] For our purpose, it is important to realise that ε presumably increases with increasing departure from the native conformation and that χ, which drives macroscopic phase separation, is temperature dependent.

‡ Note that χ and ε are strictly speaking not enthalpies but free energies because they contain information about the solvent and protein degrees of freedom that have been glossed over in our coarse-grained model.

3 A first-order conformational phase transition

The equilibrium state of the protein dispersion is given by its minimum free energy state, hence the equilibrium fraction of proteins that are in the non-native state, η_{eq}, follows by setting the partial derivative of the free energy, f_{2s} as given by eqn (1), (2) and (3), with respect to η, $\partial_\eta f_{2s} \equiv \partial f_{2s}/\partial\eta = 0$, giving,

$$\frac{\eta_{\text{eq}}}{1-\eta_{\text{eq}}} = \exp\left(\chi\phi\eta_{\text{eq}} - \varepsilon\right). \tag{4}$$

Note that we need not include contributions from steric interactions between the proteins for we presume them to have equal volume in both conformational states. An important consequence of the functional form of eqn (4) is that it, as we shall see, leads to non-universal phase behaviour, that is, there is no law of corresponding states.§

We read of from eqn (4) that the non-native state can only be stabilised by attractive interactions between proteins in the non-native state, because $\varepsilon > 0$ is a free energy penalty. For attractive interactions $\chi > 0$, stabilisation of the non-native state is only effective at sufficiently high concentrations, ϕ. This can be rationalised by realising that entropy favours a dispersion state with an equal number of proteins in each conformation, but that the free energy cost associated with the non-native state favours as few proteins as possible in that state, while attractive interactions between proteins in the non-native state favours proteins to be in the non-native state. The competition between the latter two render the value of η_{eq} strongly concentration dependent.

It is instructive to consider the actual shape of the curves given by eqn (4) for small and large values of the energetic parameters, χ and ε. The dependence of the equilibrium fraction of proteins that are in the non-native state, η_{eq}, on the protein concentration, ϕ, is shown in Fig. 2 for the two representative cases with $\chi = 3.5$ and $\varepsilon = 1$ (Fig. 2a) and $\chi = 10$ and $\varepsilon = 3$ (Fig. 2a) respectively. In the figure the solid line indicates a local free energy minimum ($\partial_{\eta\eta} f_{2s} > 0$) at fixed value of ϕ while the dashed line indicates a local maximum ($\partial_{\eta\eta} f_{2s} < 0$).

Fig. 2 shows that the dependence of η_{eq} on ϕ is fundamentally different for small and large values of χ and ε. For small values of the energetic parameters the fraction of proteins in the non-native state, η_{eq}, increases monotonically with concentration, ϕ. However, for large values of these parameters we obtain a van der Waals-like loop, indicative of thermodynamic instability. It is the presence of this thermodynamic instability that leads to the strong coupling of conformational changes within the proteins and the macroscopic phase behaviour that we eluded to in the introduction and explore in considerable detail in Sections 6 and 7.

The thermodynamic instability occurs when both energetic parameters have values larger than thermal energy, k_BT, for under these circumstances a dispersion with approximately equal amounts of proteins in the native and the non-native state

§ This equation itself cannot be rewritten as a universal equation in terms of reduced variables.

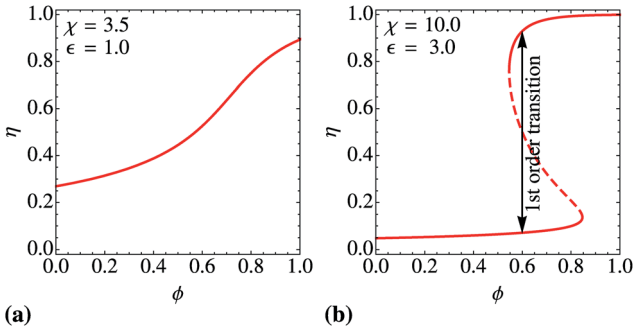

Fig. 2 The equilibrium fraction of proteins in the non-native state, η_{eq}, as a function of protein concentration, ϕ. (a) For attractive interactions of strength $\chi = 3.5k_BT$ and a free energy penalty of $\varepsilon = 1k_BT$ associated with the non-native state. (b) For $\chi = 10$ and $\varepsilon = 3$, solid lines represent stable equilibria while the dashed line represents unstable equilibria, both at fixed concentration, *i.e.*, we suppress phase separation. The location of the first-order conformational phase transition is shown.

is no longer entropically stabilised. The equilibrium state of the dispersion is then either one where nearly all proteins are in the native state or one where nearly all protein are in the non-native state. These two equilibrium states are separated by a first-order conformational phase transition at constant concentration as indicated in Fig. 2. Of course, we must realise that we cannot keep the local concentration fixed and as it turns out this phase transition is intimately coupled to macroscopic phase separation, that for now we have suppressed.

Of the dispersion states separated by the van der Waals-like loop, at any concentration one of the two states is a meta-stable state and one is the equilibrium state, and the first-order conformational phase transition occurs when both dispersion states have equal free energy density. This is illustrated in Fig. S-1,† where we show the dimensionless free energy density at $\eta = \eta_{eq}$ (shown in Fig. 2b) as a function of protein concentration for $\chi = 10$ and $\varepsilon = 3$. Not surprisingly, it shows a similar van der Waals-like loop, confirming that we are dealing with a first-order transition.

The presence of this van der Waals-like loop has an important kinetic consequence. Presuming model-C-like kinetics,[41] and having suppressed possible phase separation, the average protein conformation, η, will at all times spontaneously relax towards a value where the free energy is at a (local) minimum for the given concentration. While in the absence of the van der Waals-like loop conformational relaxation at a given concentration is always towards a unique dispersion state, this no longer holds true when the van der Waals-like loop is present. Here, the unstable part of the loop, *i.e.*, the part that corresponds to a local maximum in the free energy is a barrier separating dispersion states that relax towards a dispersion of mostly native and a dispersion of proteins mostly in their non-native state respectively. In Section 6, we shall see that this has important consequences because the thermodynamic stability of the dispersion with respect to phase separation in these two states is not necessarily equal.

The first-order conformational phase transition between dispersion states exists only for $\chi \geq 4$ and $\varepsilon \geq 2$, and in fact this demarcates the regimes of weak and strong coupling. In Section S-2,† we show more precisely under what circumstances the van der Waals-like loop and the first-order transition occur and in Section 6 we show how they couple to macroscopic phase separation. But before that, we need to formulate the contribution to the free energy that take into account volume-exclusion between the proteins. For this purpose we rely on a simplistic lattice fluid model[38,39] as well as on the more accurate Carnahan–Starling equation of state.[40]

4 Solution model

The simplest model that one can write down for volume exclusion and mixing in a binary fluid is the Flory–Huggins model.[38,39] However, this presumes protein and solvent molecules to be roughly equally sized. Obviously this is not the case for a protein solution. On the other hand, water is a structured fluid implying that plausibly one can model water on the level of clusters.[42] A somewhat more sophisticated treatment is based on the Carnahan–Starling equation of state for a hard-sphere fluid,[40] which we also consider. As we shall see, both predict qualitatively identical phase behaviour, showing that our results are robust.

At the level of a Flory–Huggins-type lattice fluid model, we have the following contribution to the free energy resulting from mixing and volume exclusions,[38,39]

$$f_{FH} = \phi \ln \phi + (1 - \phi)\ln 1 - \phi, \tag{5}$$

where f is the dimensionless free energy density, scaled to the volume of a single protein, v_0, and in units of thermal energy, k_BT and ϕ is the protein volume fraction. Within the Carnahan–Starling treatment,[40] this free energy density reads,

$$f_{CS} = \phi\left[\ln \phi - 1 + \frac{\phi(4 - 3\phi)}{(1 - \phi)^2}\right]. \tag{6}$$

We can now combine the latter two free energy densities with that of our two-state model, (3). This gives the total dimensionless free energy density of the model protein solution for the two models,

$$f_{2sFH} = f_{FH} + f_{2s}, \text{ and } f_{2sCS} = f_{CS} + f_{2s}. \tag{7}$$

From these free energy densities we can calculate the binodals and spinodals, and hence construct phase and stability diagrams. In our case, the binodal describes coexistence between protein-rich and protein-poor phases whilst the spinodal demarcates the limit of thermodynamic stability of a homogeneous dispersion.

The former can be calculated by setting temperature, osmotic pressure and the chemical potential of the proteins equal in both coexisting phases. Equal temperature implies equal χ and ε in both phases, equal chemical potential of the proteins in the two phases implies equal value of $\mu = \partial_\phi f$, and $\partial_\eta f = 0$ for both phases.¶ Finally, equal osmotic pressures

¶ Note that η is a non-conserved order parameter and hence its chemical potential must be zero.

implies equal value of $f - \phi\mu$. The resulting set of equations must be solved numerically as a function of χ and ε to determine the composition of the coexisting phases.

The spinodal lines are given by lines of inflection on the free energy surface. Such lines are obtained by setting the determinant of the Hessian equal to zero, so $\partial_{\phi\phi}f\partial_{\eta\eta}f - (\partial_{\phi\eta}f)^2 = 0$, which can be solved analytically. The inflection lines are a complex set of lines in the ϕ–η plane and are shown for the Flory–Huggins-based free energy as green dotted lines in Fig. 3, in the area enclosed by these lines the dispersion is thermodynamically unstable against macroscopic phase separation. In Fig. 3, we again compare typical cases of weak and strong coupling and superimpose the spinodal region over the equilibrium value of the fraction of proteins in the non-native state, η_{eq}, as a function of the concentration, ϕ.

We find that in the strong-coupling regime, where the van der Waals-like loop is observed, the unstable part of the loop is at all times located within the spinodal region. This shows, as advertised, that the first-order conformational phase transition and macroscopic phase separation are intimately coupled. This becomes even more apparent if we consider the binodal points, which are indicated in the same figure as blue squares and are separated by the underlying *conformational* phase transition in the strong-coupling regime. The local maximum in the free energy, indicated by the dashed line, is presumably a barrier that effectively separates the coexisting phases and might give rise to nucleation phenomena inside of the spinodal region depending on the initial conformational state of the proteins.

If we instantaneously prepare our system in a non-equilibrium state in the spinodal region, then the subsequent manner in which the protein dispersion relaxes should depend on the average conformation of the proteins prior to the quench, the free energy landscape and on the ratio, r_t, of the time scale at which conformational changes of a protein can take place and the self-diffusion time of the proteins. The latter follows from an analysis of the non-equilibrium behaviour in terms of a set of model C like kinetic equations.[41] While a complete analysis of the non-equilibrium behaviour is outside of the scope of this paper, borrowing notions from kinetic theory does allow us to analyse certain aspects of it and pinpoint in the phase diagram what kind of kinetics predominates: nucleation and growth, spinodal decomposition or a combination of both. For our system, this turns out to be highly complex, non-universal and dependent on ϕ, η, χ and ε.

For this, it makes sense to consider two limiting cases for r_t. In the first, $r_t \to \infty$, and the proteins are conformationally frozen, that is, η does not change on the diffusional time scale. The dispersion is effectively a three component dispersion, which perhaps is less interesting. In the second, more interesting case where $r_t \to 0$, relaxation of the protein conformation is instantaneous on the time scale of diffusion. In this situation, the states available to the protein dispersion are in effect restricted to states where $\eta \equiv \eta_{eq}$ and the stability of the dispersion is determined by the "equilibrium" spinodal points which are defined as the intersection of the spinodal lines and the η_{eq} curve and are shown as green dots in Fig. 3.

Presuming the limit of $r_t \to 0$ to hold we are now able to indicate various kinetic regimes in phase diagrams. This we do in the next two sections, where we demonstrate that for both thermodynamic solution models an identical demarcation between regimes of weak and strong coupling between conformational changes and macroscopic phase separation exists. In Section 7 we briefly discuss how a finite non-zero value of r_t affects the results that we present in the next two sections and show that this does not invalidate these results.

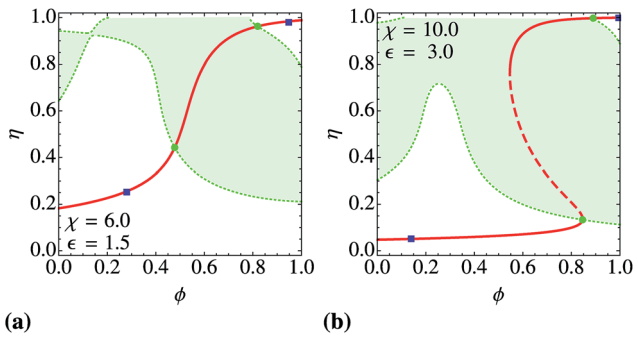

Fig. 3 (a) η_{eq} as a function of ϕ for $\chi = 6$ and $\varepsilon = 1.5$ (red line, see Fig. 4 for details), coexisting states (blue squares) and the thermodynamically unstable region (shaded light green) enclosed by spinodal lines (green dotted line). The intersection of the spinodal line and the η_{eq} curve (green dot) is defined as the "equilibrium" spinodal points. (b) The same plot for $\chi = 10$, $\varepsilon = 3$. Both plots are for the Flory–Huggins-type free energy, plots for the Carnahan–Starling-based free energy are shown in Fig. S-2.†

5 Phase diagrams for the weak-coupling regime

In this section, we present phase and stability diagrams by taking $\varepsilon = 0$ and $\varepsilon = 2$, *i.e.*, focus on the weak-coupling regime first that applies if $\varepsilon \leq 2$. We project the resulting phase and stability diagrams onto the ϕ–χ plane and retrieve Flory–Huggins-like phase behaviour,[38,39] where changes in protein conformation are induced by macroscopic phase separation. Striking differences, however, are the loss of universality and the emergence of re-entrance in the phase behaviour. These differences disappear in the hypothetical limit $\varepsilon \to -\infty$ where in our model the native state has vanishing probability.

The phase and stability diagram for the Flory–Huggins-based model for $\varepsilon = 0$ is shown in Fig. 4a, and that for $\varepsilon = 2$ in Fig. 4b. Those for the Carnahan–Starling-based model and the same values of ε in Fig. 4c and d. The binodal is represented by the blue line, where coexisting states are joined by horizontal tie lines. The collection of "equilibrium" spinodal points, as defined in the previous section, is represented by the green dotted line. Each of the diagrams consists of three regions. As is the case for the standard Flory–Huggins diagram, in region I the equilibrium state of the dispersion is a homogeneous state, in region II macroscopic phase separation occurs by nucleation and growth, while in region III it occurs by spinodal decomposition.

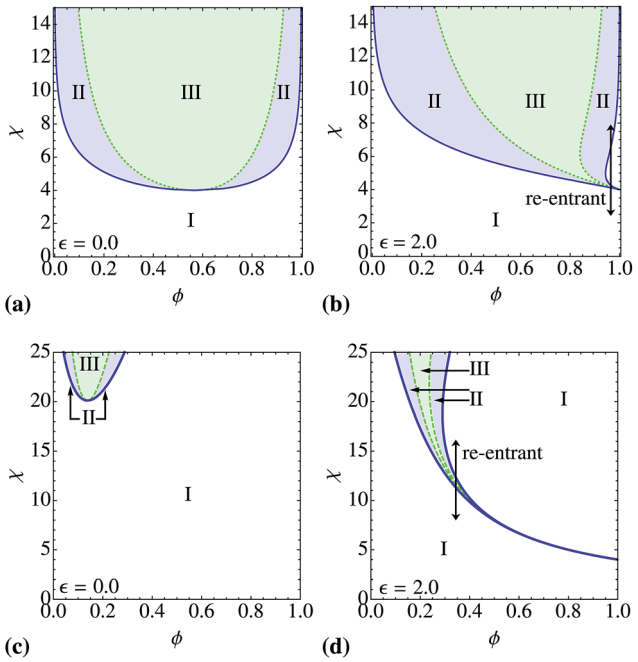

Fig. 4 Phase and stability diagrams for our two-state protein model as a function of the interaction parameter χ and protein volume fraction ϕ. (a) Results from Flory–Huggins-based model for free energy difference between native and non-native states of $\varepsilon = 0$ (units of thermal energy). (b) Flory–Huggins-based model for $\varepsilon = 2$. (c) Carnahan–Starling-based model for $\varepsilon = 0$. (d) Carnahan–Starling-based model for $\varepsilon = 2$. The binodal is depicted as a blue line, the spinodal is given by the dotted green line. In region I the equilibrium state is homogeneous, in region II phase separation occurs by nucleation and growth, in region III by spinodal decomposition. 3D versions, including the η-direction, of these diagrams can be found in Fig. S-3.†

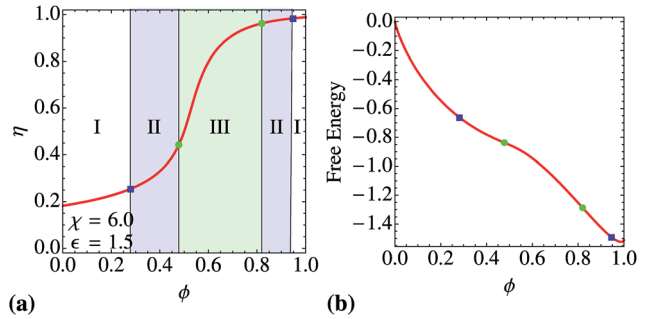

Fig. 5 (a) Equilibrium fraction of proteins in the non-native state, η_{eq}, as a function of ϕ for $\chi = 6$ and $\varepsilon = 1.5$, showing "equilibrium" spinodal points (green dots) and binodal points (blue squares), calculated from the Flory–Huggins-based model. Regions I, II and III are defined as in Fig. 4. (b) Dimensionless Flory–Huggins-based free energy density along the η_{eq} curve, as shown in (b), including the indicated binodal and "equilibrium" spinodal points. Identical plots for the Carnahan–Starling-based free energy are shown in Fig. S-4.†

By invoking the previously defined limit of $r_r \to \infty$, in which conformational changes of the protein occur instantaneously on the time scale of the self-diffusion of the proteins, we can address an important question: how does the average conformation of the proteins, as expressed in the value of η for a given concentration ϕ, affect the pathway towards thermodynamic equilibrium and why do we consider the coupling to be weak for $\varepsilon \leq 2$? To answer this question, we consider the relaxation of a protein solution as a function of the concentration, ϕ, and the protein conformation, η, for the specific case of an interaction strength $\chi = 6$ and free energy penalty on the non-native state $\varepsilon = 1.5$, and refer to Fig. 5a and b. This case is representative of the behaviour in the weak coupling regime.

In Fig. 5a, the equilibrium fraction of proteins in the non-native state, η_{eq}, is given as a function of protein concentration, ϕ. The green dots represent the "equilibrium" spinodal points and the blue squares coexisting states. Also drawn are the regions in the ϕ–η plane, where the equilibrium state of the dispersion is a homogeneous dispersion (region I) and where macroscopic phase separation occurs by nucleation and growth (region II, blue) and by spinodal decomposition (region III, green). The boundaries between these regions are given by vertical lines, so at fixed concentration, through the binodal and the "equilibrium" spinodal points. This indeed shows that the predicted kinetic mechanism by which phase separation occurs is independent of the average protein conformation, η. This must be so because of our presumption that conformational relaxation of the proteins occurs instantaneously on the diffusion time scale. In other words, any solution state first relaxes to $\eta \equiv \eta_{eq}$ at fixed concentration before diffusion can cause local changes in concentration and hence the kinetic mechanism by which phase separation occurs is independent of protein conformation, η.

Effectively, the states available to the dispersion are restricted to states where $\eta \equiv \eta_{eq}$, and as a consequence a local change in concentration, ϕ, induces a corresponding change in η, such that the corresponding state lies on the η_{eq} curve. Hence, changes in protein conformation are enslaved by changes in local concentration. For these reasons, we consider the coupling between conformational changes and macroscopic phase behaviour weak. We shall see that this does not hold true for $\varepsilon > 2$, where the coupling is strong.

Before discussing the strong-coupling regime, we consider how the phase and stability diagram changes when ε increases from 0 to 2. For both thermodynamic models (Flory–Huggins and Carnahan–Starling) the critical point shifts to larger concentrations with increasing value of ε. The reason for this is simple, the increasing conformational free energy cost of the transition to the non-native state must be compensated for by an increasing number of contacts between proteins in the non-native state. Interestingly, the phase behaviour exhibits (double) re-entrance for $\varepsilon = 2$. As indicated by arrows in Fig. 4b and d, within a limited concentration range, a continuous change in temperature and hence in χ causes the equilibrium state to shift from homogeneous to phase separated, to homogeneous and back to phase separated again. Re-entrance occurs approximately only for $\varepsilon > 1.8$ for the Flory–Huggins-based model and for $\varepsilon > 1.65$ for the Carnahan–Starling-based model.

For $\varepsilon = 2$, the critical point lies for both solution models at the (admittedly unphysical) volume fraction of $\phi = 1$ and turns out to represent a multi-critical point. At that point, the critical

point of liquid–liquid phase separation coincides with the critical point of an Ising-like conformational demixing transition. The latter gives rise to spatial domains characterised by different average conformation of the proteins at a fixed volume fraction of $\phi = 1$. This is to be expected, because the free energy density for $\phi = 1$ reduces to that of a mean-field Ising model. Note that at the critical point, for $\varepsilon = 2$ the binodal has the shape of a cusp. Interestingly, as we shall see in the next section, for $\varepsilon > 2$ the second order Ising-like transition turns into a first-order transition and there is no longer a real critical point for liquid–liquid phase separation. Finally we note that the diagrams shown in Fig. 4a–d cannot be rescaled by the critical concentration and χ value to obtain a universal phase diagram, that is, the phase behaviour is non-universal and there is no law of corresponding states.

6 Phase diagrams for the strong-coupling regime

The cross-over to strong coupling occurs at $\varepsilon = 2$ and coincides with the presence of the multi-critical point at $\phi = 1$ and $\chi = 4$: it is the convolution of the critical point of liquid–liquid phase separation and that associated with the Ising-like conformational demixing at $\phi = 1$ as discussed in the previous section. The latter transition also corresponds to the first appearance of the first-order conformational phase transition and the associated van der Waals-like loop discussed in Section 3. It is the presence of the van der Waals-like loop, in particular the unstable part of it, that gives rise to the strong-coupling regime. In the limit of $r_t \to 0$, in which conformational relaxation of the protein occurs instantaneously as discussed in Section 4, this part of the loop is a barrier separating non-equilibrium dispersion states that instantaneously relax towards a dispersion state on either the lower or upper stable part of the loop. As we shall see, the kinetic mechanism by which phase separation subsequently occurs depends onto which of the two branches the system initially relaxes.

In Fig. 6 the phase and stability diagram for the Flory–Huggins-based free energy is shown for $\varepsilon = 3$. In it, the diagram is projected onto the ϕ–χ plane. The corresponding diagram for the Carnahan–Starling-based free energy is shown in the ESI, Fig. S-5.† As before, the blue line represents the binodal and coexisting states are joined by horizontal tie lines, the green dotted line indicates the spinodal, while the region in between the red lines, which are situated mostly underneath the spinodal, demarcate the region where the van der Waals-like loop is observed. No fewer than a total of 13 regions, instead of just 3 for $\varepsilon \leq 2$, have been identified in the diagram. Three of these regions (XI–XIII) are situated between the spinodal line and the upper boundary of the region where the first-order conformational phase transition exists. These regions are small in the diagram, and hence for clarity their location is shown schematically at the bottom of the phase diagram. The background colours indicate the type of phase behaviour observed in each of the regions.

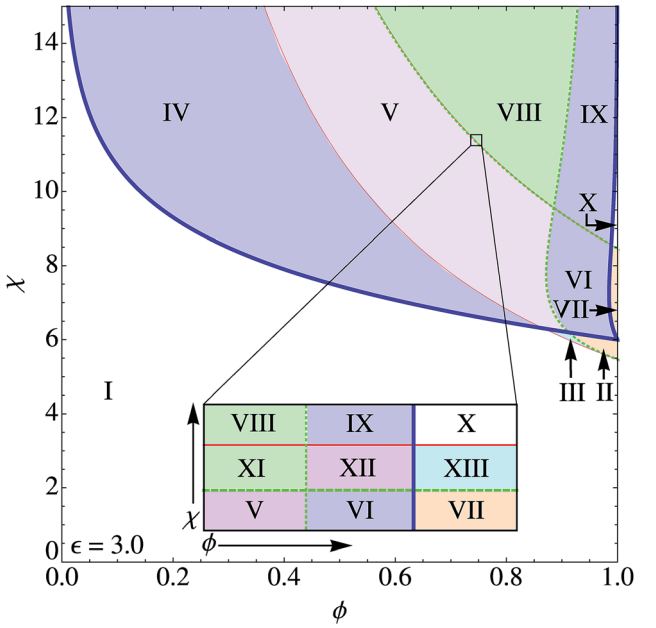

Fig. 6 The phase and stability diagram for $\varepsilon = 3$ for the Flory–Huggins-based free energy. The thick blue line represents the binodal, coexisting states are joined by horizontal tie-lines, the green dotted line represents the spinodal and the area enclosed by the red lines indicates the region where the van der Waals-like loop is present. Thirteen regions, each with distinct phase behaviour, as discussed below, are indicated. A 3D version of the diagram is shown in Fig. S-6a† and the diagram for the Carnahan–Starling-based free energy is shown in Fig. S-5.†

Before we address the phase behaviour in each of these regions, a few general remarks are in order. Firstly, the diagram as shown in Fig. 6 is a good representation of the phase behaviour for $\varepsilon \geq 2$ of both solution models, see Fig. S-6 and S-7.† Hence in the following discussion we focus entirely on the behaviour of the Flory–Huggins-based model for $\varepsilon = 3$. Before we discuss the phase behaviour in the strong coupling regime it is worth repeating that it is the unstable part of the van der Waals-like loop that separates non-equilibrium dispersion states that instantaneously relax towards either the lower (a) or upper (b) stable part of the loop that is responsible for the complexity of the diagram. This becomes evident by taking ϕ–η slices out of the phase and stability diagram and including the η_{eq} curve for $\chi = 5.95$ (Fig. 7) and $\chi = 10$ (Fig. 8), for $\chi = 7$ (Fig. S-8),† $\chi = 8.5$ (Fig. S-9)† and $\chi = 9$ (Fig. S-10),† all for $\varepsilon = 3$. We discuss the first two of these figures next.

All regions shown in Fig. 7a are part of regions I–III in Fig. 6. Just as in the weak coupling regime, boundaries between these regions are at fixed value of ϕ, *i.e.*, independent of η. Unlike in the weak coupling regime, there is an additional subdivision of regions II and III into two parts, which are separated by the unstable part of the van der Waals-like loop.

In region I, the dispersion is homogeneous and the equilibrium dispersion state is independent of the initial (non-equilibrium) value of η. In principle, the same holds true for regions II and III, however, here the presence of the van der Waals-like loop leads to the existence of meta-stable dispersion

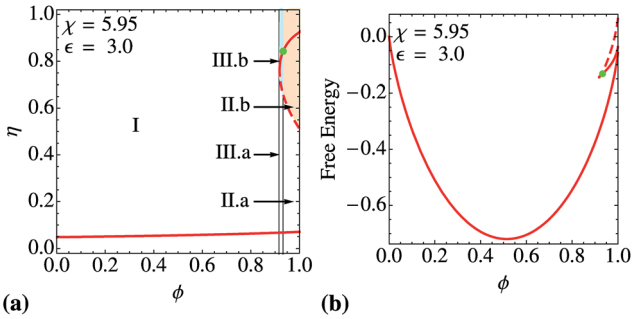

Fig. 7 (a) Mapping of phase behaviour in $\phi-\eta$ space for $\chi = 5.95$ and $\varepsilon = 3.0$ for the Flory–Huggins-based free energy, showing the η_{eq} curve (red), "equilibrium" spinodal point (green dot). Regions I, II and III are situated in the corresponding regions in Fig. 6, the behaviour in each of these regions is discussed below. (b) The dimensionless Flory–Huggins-based free energy density along the η_{eq} curve for $\chi = 5.95$ and $\varepsilon = 3.0$.

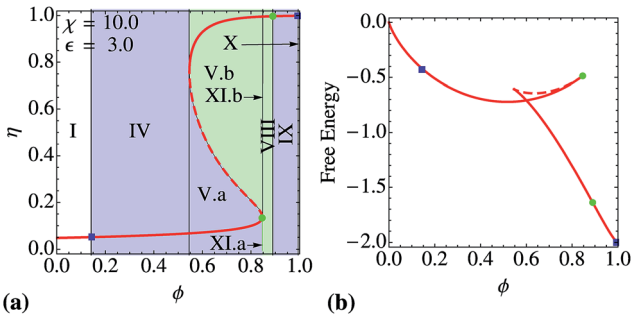

Fig. 8 (a) Mapping of phase behaviour in $\phi-\eta$ space for $\chi = 10.0$ and $\varepsilon = 3.0$ for the Flory–Huggins-based free energy, showing the η_{eq} curve (red), "equilibrium" spinodal points (green dots) and binodal points (blue squares). The regions as indicated are situated in the corresponding regions in Fig. 6, the behaviour in each of these regions is discussed below. (b) The dimensionless free energy density along the η_{eq} curve for $\chi = 10.0$ and $\varepsilon = 3.0$.

states that are located on the upper stable part of the loop. This becomes clear from Fig. 7b in which we show the free energy as a function of ϕ. In region II.b these meta-stable states are thermodynamically stable against density fluctuations, while in region III.b these meta-stable states are thermodynamically unstable because the "equilibrium" spinodal has been crossed. If the activated relaxation towards the equilibrium state on the lower stable branch is sufficiently slow, one might expect to observe a spinodal decomposition-like process to occur on the upper branch in this region. This is surprising as there are no coexisting phases and the equilibrium state is a homogeneous dispersion with most proteins in their native state. The behaviour of the model dispersion in regions VII and XIII is similar to that in respectively regions II and III and is discussed in detail in the ESI, Fig. S-9.†

Focusing now on the case $\chi = 10$ and $\varepsilon = 3$, phase separation is possible and the exact phase behaviour of the dispersion depends on both ϕ and η. This is shown in Fig. 8a, in which a total of 7 different regions are shown. As before, in regions I and X the equilibrium state of the dispersion is homogeneous and independent of the initial value of η, see Fig. 8b. In regions IV and IX phase separation must occur by nucleation and growth and in region VIII by spinodal decomposition. In regions V and XI the kinetic pathway by which phase separation occurs depends on the initial average protein conformation, η. In region V.a phase separation proceeds by nucleation and growth and by spinodal decomposition in region V.b. The behaviour in regions VI and XII is similar to that in regions V and XI with phase separation by nucleation and growth or spinodal decomposition depending on the initial average protein conformation, see the ESI Fig. S-7–S-9.† A short summary of the phase behaviour in the thirteen regions is given in Table S-1.†

An interesting feature of the phase behaviour of our model proteins in the strong-coupling regime is that, *if* initially most proteins are in their native state, the concentration interval where phase separation occurs by spinodal decomposition is small. In fact, the only region where this occurs are regions XI.a and XII.a. It arguably does not happen in region VIII because here spinodal decomposition must be preceded by relaxation of nearly all proteins to their non-native state, which plausibly does not occur instantaneously. The concentration interval over which we expect to see spinodal decomposition, $\Delta\phi_s$, is plotted as a function of χ and ε for both the Flory–Huggins and Carnahan–Starling-based free energy in Fig. 9. In that figure, and also in Fig. 6, we see that for sufficiently small values of χ the regions XI and XII are not present at all, and phase separation must always be nucleated in a dispersion of proteins in their native state.

In conclusion, in the strong coupling regime phase separation in a dispersion of mostly native proteins must almost always occur by nucleation and growth. It must be remarked that this conclusion follows from a somewhat limited kinetic analysis that is based on the presumed limit of instantaneous conformational relaxation, $r_t \to 0$. In spite of this, it is clear that the unstable part of the van der Waals-like loop poses a barrier that effectively demands the model dispersion to phase separate through nucleation and growth at least if most proteins are initially in their native state.

7 Discussion & conclusions

We presented a theoretical study into the effects of the coupling of conformational changes and protein phase behaviour. In our model, native state proteins can reversibly switch to a high-

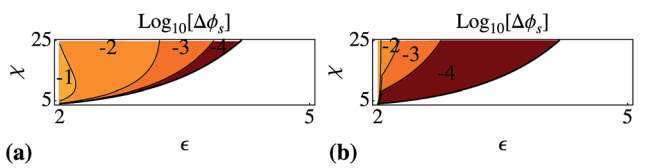

Fig. 9 Size of the concentration interval, $\Delta\phi_s$, for which phase separation occurs by spinodal decomposition if the proteins are initially in their native state as a function of χ and ε (as defined in Fig. 1). For (a) the Flory–Huggins-based and (b) the Carnahan–Starling-based free energy. Note that in the non-shaded region phase separation never occurs by spinodal decomposition.

energy non-native (*e.g.*, unfolded) state. Both conformers interact through volume exclusion, whilst proteins in the non-native state also attract each other. For simplicity we assumed that both conformers are of equal shape and volume. The model allows us to study the phase behaviour as a function of the energy difference between the native and non-native state and the strength of the interaction between proteins in their non-native state.

Our results show that there are two regimes, a regime of weak and one of strong coupling, which are demarcated by a free energy difference of $2k_BT$ between the two conformers. In the first regime, so for small free energy differences, the coupling between conformational changes and phase behaviour is weak. The phase behaviour is reminiscent of classical phase separation between solvent and solute. However, there are important differences: (1) the phase behaviour is non-universal, (2) under the right conditions the dispersion exhibits double re-entrance, and (3) changes in concentration induce changes in protein conformation.

The strong coupling regime manifests itself in three clear ways: (1) the classical critical point for phase separation disappears, (2) there is a conformational phase separation in the solvent-free (dry) system, (3) the emergence of unusual meta-stable dispersion states. This leads to a plethora of kinetic regimes, a total of 13 for both models investigated. These regimes reflect differences in the kinetic pathway towards a phase separated dispersion, depending on the average protein conformation prior to phase separation. For example, for a given concentration, phase separation can occur by nucleation and growth if most proteins are initially in their native state, while it occurs by spinodal decomposition if this is not the case.

Significantly, it turns out that phase separation in a dispersion of native proteins must almost always proceed by nucleation and growth. In that case, phase separation must be initiated by a coherent change in the conformation of proteins. Protein conformation is no longer enslaved to concentration and hence mass transport. One of our most unusual findings is that there are regimes in which there is a kinetic pathway between homogeneous dispersion states that, according to our interpretation of the model, must involve temporary phase separation.

In current protein literature, modeling of phase separation in solution ignores possible changes in conformation,[13–15,18,20] while the modeling of aggregation into supramolecular assemblies, including amyloid fibrils, hinges on changes in conformation.[23–26] We argue that the demarcation between weak and strong coupling of phase behaviour and conformational changes as predicted by our model is reminiscent of this distinction. We realise that our model does not allow for any amyloid-like structure to appear. Furthermore, because we employ equilibrium theory any irreversible binding or conformational changes is ignored, so this conclusion is tentative.

Our analysis of the thermodynamic stability and kinetic pathways towards equilibrium relies on the presumed limit of $r_t \to 0$. This means that conformational changes occur instantaneously on the self-diffusion time scale of the proteins. This assumption is probably not always realistic, yet it simplifies the kinetic analysis. If the limit of $r_t \to 0$ does not hold, the kinetics by which phase separation occur depends at all times on the initial protein conformation. The simple subdivision in phase separation by spinodal decomposition or by nucleation and growth no longer holds. A more subtle processes is possible, where at first phase separation must occur by nucleation and growth, but where, before a nucleation event has occurred, conformational relaxation forces the dispersion into a thermodynamically unstable state and phase separation proceeds by spinodal decomposition. We aim to study this and similar kinetic processes in more detail in future work from the perspective of dynamic density functional theory.

The consequences of these more complicated kinetic pathways towards a fully phase separated dispersion leave our conclusions intact. In the weak coupling regime, conformational relaxation is, at a given concentration, always towards the same average protein conformation. However, in the strong coupling regime we have seen that, depending on the initial average conformation, the dispersion relaxes toward either of two different dispersion states. Because the thermodynamic stability of these two states is not necessarily equal, this presents a fundamentally different coupling between conformation and phase behaviour. Hence, the demarcation between the weak and strong coupling regimes holds beyond the presumed limit of $r_t \to 0$.

Much of the interesting behaviour predicted by our models occurs at very high concentration, where the protein might crystallize. It would be interesting to see how the phase behaviour predicted by our model couples to crystallization. We have invoked a simple van der Waals-solid model,[43] which led to unphysical results. As there are no attractive interactions between proteins in the native state within our model, phase separation leads to a dense phase of attractively interacting proteins in their non-native state. Phase separation is then always thermodynamically more favourable than crystallization of non-interacting native proteins. Clearly, more work is required to couple our model to a crystallization model.

Finally, a few remarks must be made on the similarities and differences between the Flory–Huggins-type and Carnahan–Starling-type free energies. Although the phase and stability diagrams appear to be distinctly different in shape for both models, the underlying structure is identical and hence both models support the existence of a demarcation between a weak and strong coupling regime. The main cause for the difference in shape of the diagrams is that in the Carnahan–Starling-type model the free energy more strongly increases with protein concentration, leading to the formation of a less dense dispersion of proteins in the non-native state by phase separation.

Acknowledgements

The authors acknowledge the Dutch Polymer Institute (DPI), TI Food Nutrition (TIFN), and Foundation for Fundamental Research on Matter (FOM) for providing the funding for this research (Project number: 10BRM47) and thank A. M. Donald, R. H. Tromp and J. C. Ioannou for discussions.

Notes and references

1 A. Syrbe, W. Bauer and H. Klostermeyer, *Int. Dairy J.*, 1998, **8**, 179.
2 E. van der Linden and P. Venema, *Curr. Opin. Colloid Interface Sci.*, 2007, **12**, 158.
3 A. M. Donald, *Soft Matter*, 2008, **4**, 1147.
4 R. Mezzenga and P. Fischer, *Rep. Prog. Phys.*, 2013, **76**, 046601.
5 E. Y. Chi, S. Krishnan, T. W. Randolph and J. F. Carpenter, *Pharm. Res.*, 2003, **20**, 1325.
6 A. Saluja and D. S. Kalonia, *Int. J. Pharm.*, 2008, **358**, 1.
7 J. C. Kendrew, H. M. D. G. Bodo, R. G. Parrish and H. Wyckoff, *Nature*, 1958, **181**, 662.
8 S. B. Zimmerman and A. P. Minton, *Annu. Rev. Biophys. Biomol. Struct.*, 1993, **22**, 27.
9 R. Ellis, *Trends Biochem. Sci.*, 2001, **26**, 597.
10 F. Chiti and C. M. Dobson, *Annu. Rev. Biochem.*, 2006, **75**, 333.
11 S. Zhang, *Nat. Biotechnol.*, 2003, **21**, 1171.
12 O. Ptitsyn and A. V. Finkelstein, *Protein Physics: A Course of Lectures*, Academic Press, 2002.
13 R. P. Sear, *J. Chem. Phys.*, 1999, **111**, 4800.
14 A. Lomakin, N. Asherie and G. B. Benedek, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9465.
15 N. Kern and D. Frenkel, *J. Chem. Phys.*, 2003, **118**, 9882.
16 E. Bianchi, R. Blaak and C. N. Likos, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6397.
17 D. E. Kuehner, H. W. Blanch and J. M. Prausnitz, *Fluid Phase Equilib.*, 1996, **116**, 140.
18 M. L. Broide, T. M. Tominc and M. D. Saxowsky, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1996, **53**, 6325.
19 M. Muschol and F. Rosenberger, *J. Chem. Phys.*, 1997, **107**, 1953.
20 P. Prinsen and T. Odijk, *J. Chem. Phys.*, 2004, **121**, 6525.
21 V. G. Taratuta, A. Holschbach, G. M. Thurston, D. Blankschtein and G. B. Benedek, *J. Phys. Chem.*, 1990, **94**, 2140.
22 V. M. O. Batista and M. A. Miller, *Phys. Rev. Lett.*, 2010, **105**, 088305.
23 C. M. Dobson, *Nature*, 2003, **426**, 884.
24 M. R. Krebs, G. L. Devlin and A. Donald, *Biophys. J.*, 2007, **92**, 1336.
25 T. R. Jahn and S. E. Radford, *Arch. Biochem. Biophys.*, 2008, **469**, 100.
26 J. E. Gillam and C. E. MacPhee, *J. Phys.: Condens. Matter*, 2013, **25**, 373101.
27 A. C. Dumetz, A. M. Chockla, E. W. Kaler and A. M. Lenhoff, *Biophys. J.*, 2008, **94**, 570.
28 S. Auer and D. Kashchiev, *Phys. Rev. Lett.*, 2010, **104**, 168105.
29 A. Aggeli, M. Bell, L. M. Carrick, C. W. G. Fishwick, R. Harding, P. J. Mawer, S. E. Radford, A. E. Strong and N. Boden, *J. Am. Chem. Soc.*, 2003, **125**, 9619.
30 G. Karlstroem, *J. Phys. Chem.*, 1985, **89**, 4962.
31 P. G. Vekilov, *J. Phys.: Condens. Matter*, 2012, **24**, 193101.
32 R. Jaenicke, *J. Biotechnol.*, 2000, **79**, 193.
33 K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hubner and D. Kern, *Nature*, 2007, **450**, 838.
34 L. Hong, D. C. Glass, J. D. Nickels, S. Perticaroli, Z. Yi, M. Tyagi, H. O'Neill, Q. Zhang, A. P. Sokolov and J. C. Smith, *Phys. Rev. Lett.*, 2013, **110**, 028104.
35 J. Xie and H. Huang, *Colloids Surf., B*, 2011, **85**, 97.
36 L. Day, J. Zhai, M. Xu, N. C. Jones, S. V. Hoffmann and T. J. Wooster, *Food Hydrocolloids*, 2014, **34**, 78.
37 J. C. Ioannou, A. M. Donald and R. H. Tromp, *Food Hydrocolloids*, 2015, **46**, 216.
38 P. J. Flory, *J. Chem. Phys.*, 1942, **10**, 51.
39 M. L. Huggins, *J. Phys. Chem.*, 1942, **46**, 151.
40 N. F. Carnahan and K. E. Starling, *J. Chem. Phys.*, 1969, **51**, 635.
41 P. Hohenberg and B. Halperin, *Rev. Mod. Phys.*, 1977, **49**, 435.
42 R. Ludwig, *Angew. Chem., Int. Ed.*, 2001, **40**, 1808.
43 A. Daanoun, C. F. Tejero and M. Baus, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1994, **50**, 2913.