Soft Matter

PAPER

View Article Online
View Journal | View Issue

Cite this: *Soft Matter*, 2016, 12, 7289

Modeling phase transitions in mixtures of β–γ lens crystallins†

Miha Kastelic,[a] Yurij V. Kalyuzhnyi[b] and Vojko Vlachy*[a]

We analyze the experimentally determined phase diagram of a γD–βB1 crystallin mixture. Proteins are described as dumbbells decorated with attractive sites to allow inter-particle interaction. We use thermodynamic perturbation theory to calculate the free energy of such mixtures and, by applying equilibrium conditions, also the compositions and concentrations of the co-existing phases. Initially we fit the T_{cloud} *versus* packing fraction η measurements for a pure (x_2 = 0) γD solution in 0.1 M phosphate buffer at pH = 7.0. Another piece of experimental data, used to fix the model parameters, is the isotherm x_2 *vs.* η at T = 268.5 K, at the same pH and salt content. We use the conventional Lorentz–Berthelot mixing rules to describe cross interactions. This enables us to determine: (i) model parameters for pure βB1 crystallin protein and to calculate; (ii) complete equilibrium surface (T_{cloud}–x_2–η) for the crystallin mixtures. (iii) We present the results for several isotherms, including the tie-lines, as also the temperature-packing fraction curves. Good agreement with the available experimental data is obtained. An interesting result of these calculations is evidence of the coexistence of three phases. This domain appears for the region of temperatures just out of the experimental range studied so far. The input parameters, leading good description of experimental data, revealed a large difference between the numbers of the attractive sites for γD and βB1 proteins. This interesting result may be related to the fact that γD has a more than nine times smaller quadrupole moment than its partner in the mixture.

Received 1st July 2016,
Accepted 1st August 2016

DOI: 10.1039/c6sm01513a

www.rsc.org/softmatter

1 Introduction

Much of biology depends on proteins interacting with each other – pairwise or in form of oligomers – mediated by solvent and different ligands, such as salts and excipients.[1,2] We need to understand how protein molecules interact with water, ions and in-between them. The self-assembly of proteins into various structures plays a crucial role in biology. A better understanding of these processes is expected to be reached by modern computer simulations, which have become indispensable for insights into physics and chemistry. Yet, the systems of several protein species in water with salts and other co-solvents, which are relevant to biology and pharmacy, are often too big to be handled well enough by current explicit-water molecular simulations. A somewhat different, less detailed, approach, based on the ideas of condensed matter physics has been forwarded in recent decades. Interactions of mixtures of oppositely charged proteins have been studied experimentally[3] and theoretically.[4] Important previous studies relevant to our work are listed as ref. 5 and 6. Two excellent reviews of developments in this area of research have been published recently.[7,8]

Proteins may self-assemble to form dense liquid phase, crystals, or amorphous aggregates, and an understanding of this process is important for many reasons.[2,9–11] First, the self-assembly is a step in protein crystallization; good quality crystals are needed to determine protein structure through diffraction experiments. The latter pathway is the following: soluble protein aggregates initialize the coexistence of low- and high density liquid phases, differing in protein concentration. The so-called liquid–liquid phase separation is a meta-stable stage and may further lead to crystallization. For this reason, many experimental and theoretical efforts, were invested to improve understanding of the liquid–liquid phase separation. The pioneering work of George and Wilson,[12] and the later study of Vliegenthart and Lekkerkerker[13] revealed a strong correlation of the liquid–liquid phase boundary with the second virial coefficient and thus inter-protein interactions. Secondly, protein self-assembly in living cells plays a key role in diseases, such as Alzheimer's, Parkinson's, and others.[1,14,15] Thirdly, a key step in developing biotech drugs[16] is to formulate proteins that do not form aggregates. The challenge is to deliver formulations with high concentrations of solutes, low viscosity and maintaining the protein's therapeutics properties for at least two years. Self-assembly decreases therapeutic efficacy and increases its viscosity.

[a] *Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, 1000 Ljubljana, Slovenia. E-mail: vojko.vlachy@fkkt.uni-lj.si*
[b] *Institute for Condensed Matter Physics, Svientsitskii 1, 79011 Lviv, Ukraine*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6sm01513a

As atomistic level molecular simulations are not practical for studying phase separations in such solutions, the traditional approach is to use the colloidal model of Derjaguin–Landau–Verwey–Overbeek (DLVO).[17] In this approach, proteins are treated as spheres that interact *via* van der Waals and screened Coulomb potentials, acting between centers of molecules. The approach proved to be unsatisfactory in several respects.[1] On the other hand, many aspects of protein aggregation may be pictured by the models that include anisotropic protein–protein interactions of short range.[7,8,18–21] This strategy was successfully used to model a wide class of globular proteins and to analyze the liquid–liquid phase separation,[22,23] crystallization,[24,25] osmotic pressures,[26] distribution of aggregates,[27,28] percolation threshold,[27] and other physico-chemical properties.[21]

Heras and co-workers[29,30] extended these ideas to binary mixtures of equally-sized spherical particles. They examined an influence of pair potential characteristics on the phase behavior under isobaric conditions. A special interest for us is the work of Roldnán-Vargas[31] and coworkers. These authors treated binary mixtures of spherical particles of different sizes and with different surface patterns. The fact that the shape of protein molecules affects the phase separation, has recently been confirmed by us. A distinct shift of liquid–liquid critical density toward lower values has been obtained in case of the associating dumbbells.[32] Such a behavior was previously observed experimentally for the system of Y-shaped antibodies.[33] For proteins in dense mixtures near the phase separation a realistic modeling of protein shape seems to be important.

In the previous studies[34,35] we applied simple models with anisotropic interactions to analyze aqueous protein solutions in the presence of low-molecular-mass salts. We treated proteins as hard spheres, decorated with square-well-energy binding sites, using Wertheim's thermodynamic perturbation theory.[36] The model provided good fits to the cloud-point temperature curves of lysozyme in buffer–salt mixtures as a function of the type and concentration of salt. The necessary condition required for such modeling to be realistic is that proteins in solution during the experiment remain in their compact form.[37,38] Unfortunately, many experimentalists often do not provide this information. In the next contribution[32] the approach above was extended to examine the thermodynamics of fluid with "molecules" represented by two fused hard spheres, again decorated by the attractive square-well sites. Interaction between these sites was of short-range and caused association between the fused-sphere particles. The model served to estimate the effects of non-spherical geometry of proteins on solution properties.

Here we continue with applications of simple models to protein systems. Going one step further, we treat a binary mixture of proteins with molecules of each of the respective species represented as dumbbells. In the first step we need to form the molecule (dumbbell) by fusing two hard spheres and further decorate it with attractive sites on the surface. These sites allow association between the same type of molecules (self-association) as also with molecules of the other protein species present in the system. Thermodynamic properties of the model mixture are obtained *via* Wertheim's thermodynamic perturbation theory.[36] We utilized the McMillan–Mayer approach, where the principal variables are concentration, chemical potential of composed solvent, and temperature. We calculated the liquid–liquid phase diagram and present the results as the volume packing fraction–composition–temperature diagram. Under such conditions the osmotic pressure follows from the equation of state. When it comes to comparison with experimental data, we chose to model a γD–βB1 crystallin mixture, thoroughly studied by Wang and co-workers.[39] Both proteins show significant deviations from the spherical shape and resemble dumbbells[40] (see Fig. 1 therein, and Fig. 6 of this paper). These proteins are of special interest: an improper functioning of γD crystallin may lead to cataracts in the human eye.[14] For this reason crystallins were intensively studied. The investigations include stability measurements,[41] influence of mutation studies[42,43] and other surface modifications,[44] and the analysis of surface residues forming inter-protein salt bridges.[45] The investigations revealed that a replacement of one single amino acid with another may modify physiological function and/or change the crystallization pathway.[46–49] Investigation of this subtle phenomenon is beyond the scope of this work. In this initial study we focus to the liquid–liquid phase equilibrium, and how the latter is governed by the interactions between two protein species. We use a simple "dumbbell" protein model, coupled with the thermodynamic perturbation theory of Wertheim,[36] to analyze the experimental data of Wang and co-workers.[39]

2 Protein molecules as dumbbells decorated by attractive sites

In the first step of the calculation we compose dumbbell-like molecules of type i with number density $\rho_i^{(p)}$ by dimerization of $\rho_i = 2\rho_i^{(p)}$ hard-spheres through the sticky sites B (see Fig. 1). The sticky BB interaction, $u_{BB}^{(ij)}(z_{BB})$, is written as:

$$\langle \exp\{-\beta u_{BB}^{(ij)}(z_{BB})\} - 1\rangle_{\Omega_i\Omega_j} = \delta_{ij} K_{BB}^{(ij)} \delta(r_{ij} - \sigma), \qquad (1)$$

where r_{ij} denotes the distance between centers of spheres of the type i and j and Ω_i,Ω_j their orientations. Furthermore, z_{BB} is the

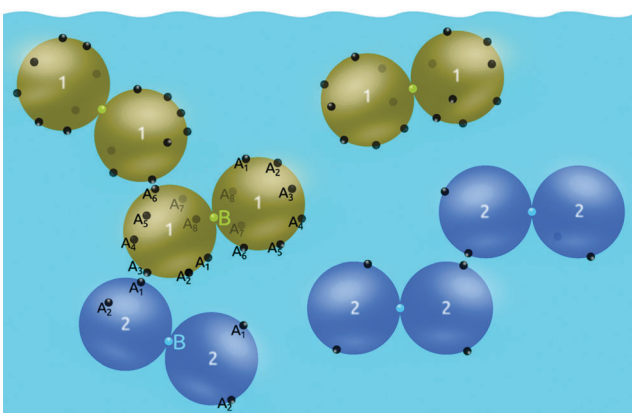

Fig. 1 Hard spheres of equal size form dumbbells (model for protein molecule) through the sticky site B. Further interaction between protein molecules is possible through the A_α sites. In this scheme $m_1 = 8$ and $m_2 = 2$.

distance between sites B, $\langle\ldots\rangle_{\Omega_1\Omega_2}$ denote the orientation average, $\delta(\ldots)$ is the Dirac delta function, and δ_{ij} is the Kroneker delta symbol. The dumbbells are formed when $K_{BB}^{(11)}$ and $K_{BB}^{(22)} \to \infty$. For simplicity we assume that the hard spheres forming the protein species are of equal size σ. In the experimental case analyzed here this is not far from being true. The number of the attractive sites of type A_α, located on a sphere belonging to the molecule i is m_i. Furthermore, α runs over 1,...,m_i, while index i is assuming values 1 and 2. The number density of the system is $\rho_t^{(p)} = \rho_1^{(p)} + \rho_2^{(p)}$.

Protein molecules i and j are interacting *via* the pair potential, which consists of the sum of four terms $u_{ij}(r_{ij},\Omega_i,\Omega_j)$. Note again that hard-spheres labeled 1 form molecules of type 1, while the hard-spheres denoted 2 form molecules of type 2. These terms describe the interaction between decorated hard spheres i and j, belonging to molecules i and j:

$$u_{ij}(r_{ij},\Omega_i,\Omega_j) = u_{hs}(r_{ij}) + \sum_{\alpha}^{m_i}\sum_{\beta}^{m_j} u_{\alpha\beta}^{(ij)}(z_{\alpha\beta}). \quad (2)$$

Here $u_{hs}(r_{ij})$ is the hard-sphere potential, $z_{\alpha\beta}$ is the distance between sites A_α and A_β, which belong to the spheres of the type i and j,

$$u_{\alpha\beta}^{(ij)}(z_{\alpha\beta}) = \begin{cases} -\varepsilon_{ij} & \text{for } z_{\alpha\beta} < \omega_{ij}, \\ 0 & \text{for } z_{\alpha\beta} \geq \omega_{ij}, \end{cases} \quad (3)$$

ε_{ij} and ω_{ij} are the square-well depth and width, respectively. In summary, we have four types of interaction, *i.e.* 1–1, 1–2, 2–1, and 2–2, between the decorated hard spheres denoted 1 and 2.

3 Thermodynamic perturbation theory

Numerical results for the model described above are obtained by Wertheim's thermodynamic perturbation theory (TPT).[36] The basic quantity, Helmholtz free energy A, is composed of the ideal (id), hard-sphere (hs), and association term (ass):

$$A = A_{id} + A_{hs} + A_{ass}. \quad (4)$$

These terms are written as:

$$\frac{\beta A_{id}}{V} = \sum_{i=1}^{2} \rho_i\left[\ln(\Lambda_i^3\rho_i) - 1\right], \quad (5)$$

$$\frac{\beta A_{hs}}{V} = \frac{4\eta - 3\eta^2}{(1-\eta)^2}\rho_t, \quad (6)$$

$$\frac{\beta A_{ass}}{V} = \sum_{i=1}^{2} \rho_i\left\{m_i\left[\ln X_i - \frac{1}{2}X_i + \frac{1}{2}\right] - \frac{1}{2}\ln\rho_i\sigma^3 g_{hs}^{(py)} + \frac{1}{2}\right\}, \quad (7)$$

where $\beta = 1/k_BT$ and T is the absolute temperature, Λ_i is the de Broglie thermal wavelength,[50] $\rho_t = \rho_1 + \rho_2$, $\eta = \pi\rho_t\sigma^3/6$, $g_{hs}^{(py)} = (1 + \eta/2)/(1 - \eta)^2$ is the Percus–Yevick expression for the contact value of the hard-sphere radial distribution function,[51] and X_i is the fraction of the particles not bonded at any site A_α of a hard sphere of type i. In addition we assume two types of steric incompatibilities: (i) the site A_α on one molecule cannot bond simultaneously to two A_α sites on another molecule; and (ii) double bonding between molecules is not allowed; see also Fig. 1 and the ref. 52 (Fig. 2 therein). X_i follows from the statistical–mechanical analogue of the mass action law:[53]

$$\sum_{j=1}^{2} m_j\rho_j\Delta_{ij}X_iX_j + X_i - 1 = 0, \quad (8)$$

where

$$\Delta_{ij} = 4\pi g_{hs}^{(py)}\int_{\sigma}^{\sigma+\omega_{ij}} \bar{f}_{ij}(r)r^2\mathrm{d}r. \quad (9)$$

Here $\bar{f}_{ij}(r)$ is the orientational average of the Mayer function for the square-well site–site interaction:

$$\bar{f}_{ij}(r) = (\exp(\beta\varepsilon_{ij}) - 1)(\omega_{ij} + \sigma - r)^2(2\omega_{ij} - \sigma + r)/(6\sigma^2 r). \quad (10)$$

Wertheim's TPT uses the orientation average of $f_{ij}(r)$ and accordingly placement of the sites does not enter the theory. There is computer evidence in the literature that this is not a severe approximation.[54]

Once the free energy A is known, the chemical potential of the protein species i, $\mu_i^{(p)}$, and the equation of state (pressure P) can be obtained from the standard thermodynamic relations[50]

$$\mu_i^{(p)} = \left[\frac{\partial(A/V)}{\partial\rho_i^{(p)}}\right]_{T,V,\rho_j^{(p)}\neq\rho_i^{(p)}}, \quad (11)$$

$$P = \sum_i \rho_i^{(p)}\mu_i^{(p)} - \frac{A}{V}. \quad (12)$$

4 Conditions of phase equilibrium

A binary system with protein densities $\{\rho_1^{(p)}(0),\rho_2^{(p)}(0)\}$ may undergo phase decomposition into Q phases, denoted here as (1), (2)...(γ)...(Q). According to the Gibbs phase rule $Q \leq 4$. The notation $\{\rho_1^{(p)}(0),\ \rho_2^{(p)}(0)\}$ denotes the (unstable) "mother" phase, while the notation $\{\rho_1^{(p)}(\gamma),\ \rho_2^{(p)}(\gamma)\}$ applies to the (stable) "daughter" phases (γ). Thermodynamic conditions to be satisfied in equilibrium read:

$$\mu_i^{(p)}(\{\rho_1^{(p)}(1),\ \rho_2^{(p)}(1)\},\ T) = \ldots = \mu_i^{(p)}(\{\rho_1^{(p)}(Q),\ \rho_2^{(p)}(Q)\},\ T), \quad (13)$$

$$P(\{\rho_1^{(p)}(1),\ \rho_2^{(p)}(1)\},\ T) = \ldots = P(\{\rho_1^{(p)}(Q),\ \rho_2^{(p)}(Q)\},\ T). \quad (14)$$

In words: the chemical potential of species i = 1,2 within phase (γ), $\mu_i^{(p)}(\gamma)$, must be at given temperature and pressure equal in all phases. In addition to these conditions, we need also to consider the conservation of the volume fractions of phases $X(\gamma) = V(\gamma)/V(0)$ and the number densities of proteins $\{\rho_1^{(p)}(\gamma),\rho_2^{(p)}(\gamma)\}$ over the coexisting phases (γ):

$$1 = \sum_{\gamma=1}^{Q} X(\gamma), \quad (15)$$

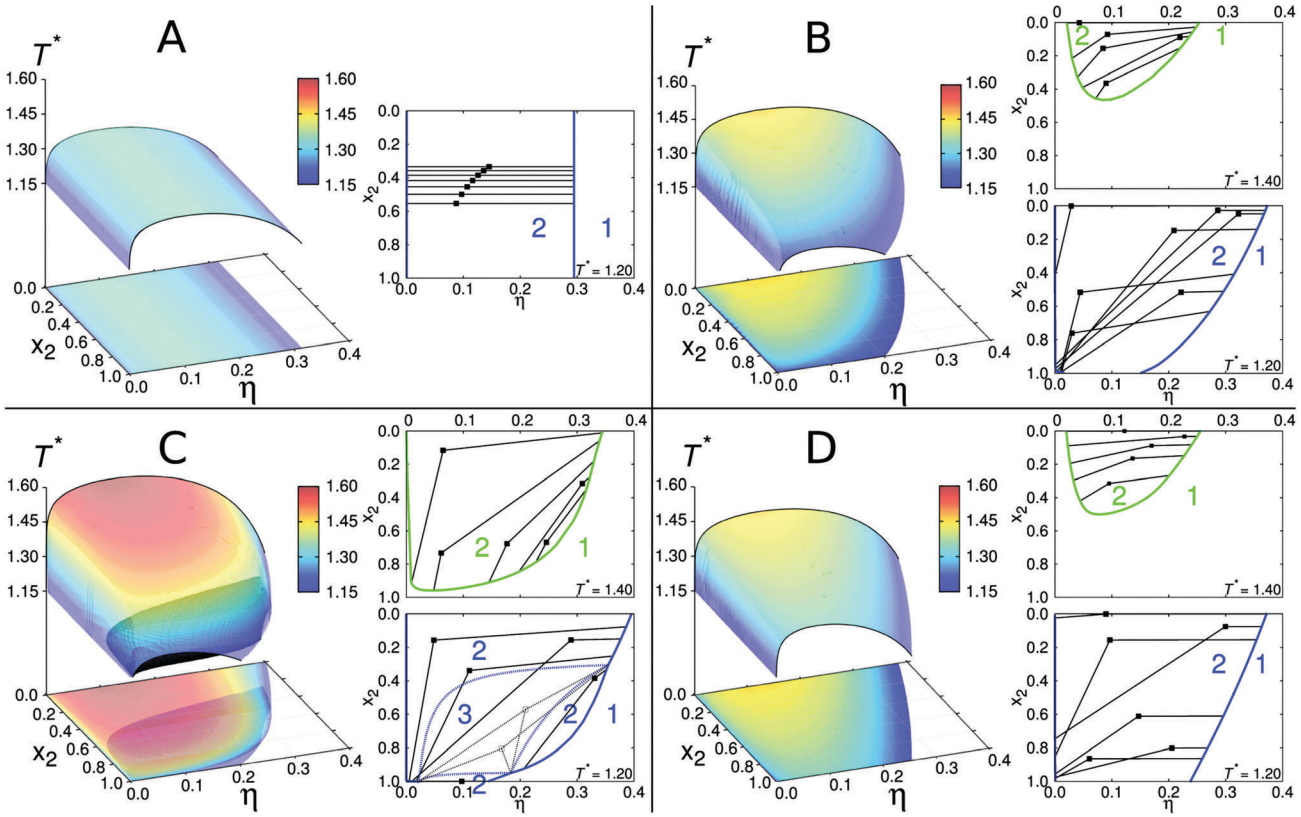

Fig. 2 Phase diagrams for binary protein mixtures. Within each panel we present a complete equilibrium surface and its projections onto the η–x_2 plane, as well as isotherms showing how the unstable points (denoted by filled squares) decompose into the coexisting liquid phases. Parameters for each panel read: $\varepsilon_{11} = \varepsilon_{22} = 13.7\varepsilon$, $m_1 = m_2 = 4$ (panel A); $\varepsilon_{11} = \varepsilon_{22} = 13.7\varepsilon$, $m_1 = 5$, $m_2 = 3$ (panel B); $\varepsilon_{11} = 1.1\varepsilon_{22}$, $\varepsilon_{22} = 13.7\varepsilon$, $m_1 = 5$, $m_2 = 3$ (panel C); and $\varepsilon_{22} = 1.1\varepsilon_{11}$, $\varepsilon_{11} = 13.7\varepsilon$, $m_1 = 5$, $m_2 = 3$ (panel D). The equality, $\omega_{11} = \omega_{22} = 0.061\sigma$, holds for all panels. The constant 13.7 was (merely for the purpose of presentation) chosen arbitrarily to cause the critical temperatures in these graphs to fall between 1.15 and 1.60.

$$\rho_i^{(\mathrm{p})}(0) = \sum_{\gamma=1}^{Q} X(\gamma)\rho_i^{(\mathrm{p})}(\gamma). \qquad (16)$$

Eqn (13)–(16) form the set of expressions for unknowns $\{\rho_1^{(\mathrm{p})}(\gamma),\ \rho_2^{(\mathrm{p})}(\gamma)\}$, $X(\gamma)$ ($\gamma = 1 \ldots Q$) at specified densities $\{\rho_1^{(\mathrm{p})}(0),\ \rho_2^{(\mathrm{p})}(0)\}$ of the mother phase.

In order to determine the coexistence region we need to solve numerically the set of equations defined above. This task may be accompanied by several problems. First, the solution of the system of nonlinear equations is not unique and the exact number of solutions is not known in advance. Second, numerical methods for solving nonlinear equations are usually based on iterative algorithms, requiring a good initial approximation to reach convergence. Third, in a process of searching for the solution one can get stranded at a saddle point, away from the global free energy minimum. In general, additional restrictions[55] must be considered, what makes such an approach time consuming.

To avoid these complications, we used a somewhat different approach, searching for the global minimum directly. The appropriate functional to be minimized is the free energy of the system per volume, $A(0)/V(0)$,

$$A(0)/V(0) = \sum_{\gamma=1}^{Q} X(\gamma) A\left(\left\{\rho_1^{(\mathrm{p})}(\gamma), \rho_2^{(\mathrm{p})}(\gamma)\right\}, T\right) \Big/ V(\gamma). \qquad (17)$$

To account for the conservation conditions given by eqn (15) and (16), we insert these expressions into eqn (17), and thus reduce the dimensionality of the problem. Next we use the Boltzmann simplex simulated annealing method (BSSA), described in Chapters 10.4 and 10.9 of ref. 56, to find the global minimum. When this minimum is reached, the conditions given by eqn (13) and (14), are automatically fulfilled. Numerically, the equations are satisfied within 10^{-2}%. More detailed description of the search for the global minimum is given as a part of the ESI.†

5 Binary mixture of model proteins

Our goal here is to study the stability of mixtures of two protein species. The main parameters of the model are: (i) number of binding sites m_1 and m_2 per sphere (for protein species 1 and 2); (ii) the attractive square well depths ε_{11}, ε_{22}; and (iii) the corresponding interaction ranges ω_{11}, ω_{22}. For the cross interactions we use standard Lorentz–Berthelot mixing rules: $(\varepsilon_{12})^2 = \varepsilon_{11}\varepsilon_{22}$ and $2\omega_{12} = \omega_{11} + \omega_{22}$. Initially we perform a systematic analysis of how main model parameters affect the phase diagrams, presented as the volume packing fraction (η) – composition ($x_2 = \rho_2^{(\mathrm{p})}/\rho_t^{(\mathrm{p})}$) – temperature ($T$) graphs.

Model parameters shape the phase diagram

The initial case to be examined is a mixture where both proteins have four sites per sphere, $m_1 = m_2 = 4$, equal square well depths $\varepsilon_{11} = \varepsilon_{22} = 13.7\varepsilon$, and corresponding ranges $\omega_{11} = \omega_{22} = 0.061\sigma$. In other words, we are mixing the same type of molecules. Note that in this chapter (and here only), we use the reduced units: energy is measured in ε, temperature as $T^* = k_B T/\varepsilon$, and length in diameter of sphere σ. Results for such a ''symmetric'' mixture are shown in panel A of Fig. 2. The states for temperatures below the equilibrium surface are unstable and decompose into low- and high- (volume) packing fraction (η) liquid phases. The ''mixture'' studied above represents the simplest possible example; it translates the symmetry of the pair potentials into topology of the phase diagram. The fact that two protein species have equal interaction parameters and sizes causes for the curves belonging different x_2 values to be parallel to each other. When the phase separation occurs, the decomposition into low- and high-packing fractions is purely density driven. This is shown on the right site of panel A (two- and one-phase regions are labeled by numbers ''2'' and ''1'', respectively). In other words, the lines connecting coexisting phases, the so-called tie lines are in this case parallel to the η axis.

To study the deviations from the basic case examined above, we place unequal numbers of sites on each protein ($m_1 \neq m_2$), while keeping other parameters unchanged. We examine the case where m_1 is increased by one ($m_1 = 5$), and m_2 decreased by one ($m_2 = 3$). The corresponding phase diagram is shown on panel B of Fig. 2. In contrast to the previous example, the x_2 curves are not parallel to each other and a distinct asymmetry along x_2 axis is observed. Envelopes for pure components, denoted by black solid lines at $x_2 = 0$ and $x_2 = 1$, are different: protein 1 has higher critical density and temperature, since it has five attractive sites (per sphere) than the protein species labeled 2. This observation is consistent with our previous study,[34] as well as, with observations of other authors studying one-component systems.[57,58] In addition to this, the unstable states exhibit higher complexity: as inferred from the right site of panel B, $T^* = 1.20$, the high-packing fraction phase tends to retain the composition of the unstable state (density driven), while the low-packing fraction phase is enriched on an account of the second component (composition driven), to satisfy the conservation of particles. At higher temperature, $T^* = 1.40$, this effect is less pronounced.

In the next example the differences between protein species are introduced on account of modifying their interaction energies ε_{ij}. It is important to stress out that overall protein–protein interaction is controlled by ε_{ij} and ω_{ij}, as well as, by the number of binding sites m_i. Here the square well depth ε_{11} is increased for 10%, while keeping the number of sites $m_1 = 5$ and $m_2 = 3$, and ranges ω_{ij} unchanged. These results are shown on panel C of Fig. 2. On the first sight there seems to be little difference to panel B, except for the higher critical point of the envelope at $x_2 = 0$. A closer look, however, reveals the coexistence region of three liquid phases (black shaded area), embedded into the two-phase region. Interestingly, the three-phase region disappears for the temperature equal to the critical value of pure protein 2 ($x_2 = 1$) that is at $T^* = 1.225$. To demonstrate the simultaneous presence of two- and three-phase regions, we on the same panel show also the isotherms for temperatures below and above $T^* = 1.225$. The isotherm for $T^* = 1.20$ reflects the presence of three-phase region, labeled by number ''3'', and surrounded by a closed dashed-blue loop. Unstable states within this region (denoted by empty squares) decompose into the phases with packing fraction and composition as indicated on the Figure. We may call these three phases the low-, intermediate-, and high-packing fraction liquid phases. According to the Gibbs phase rule, in a two-component mixture the region where three phases are in equilibrium has only one degree of freedom, in our case this is temperature. The three-phase region separates two-phase region into three sub-regions as specified on the Figure. At higher temperature, $T^* = 1.40$, a simpler phase diagram having only two phases in equilibrium is re-discovered.

Finally, in contrast with the previous example, we keep $\varepsilon_{11} = 13.7\varepsilon$ ($\varepsilon = k_B T/T^*$) unchanged and increase the ε_{22} for 10%. As before, $m_1 = 5$ and $m_2 = 3$. In this way we suppress – in comparison with the previous case – the interaction potential difference between the two protein species. The resulting phase diagram is shown in panel D of Fig. 2. In this case the phase behavior exhibits a higher symmetry along the x_2 axis than in panels B and C: the shape of the envelope at $x_2 = 1.0$ resembles the one at $x_2 = 0.0$. Interestingly, the three-phase region shown in panel C disappears under such conditions. Because of similarity in interactions among the two protein species, the system does not need a third phase for stabilization. Furthermore, the high packing fraction (daughter) phases, derived from unstable states (mother phases), tend to retain the composition of the mother phases, similarly as before shown on panel B.

6 Analysis of experimental data for γD–βB1 mixtures

Encouraged by these results we turned to analyze experimental data for γD–βB1 protein mixtures, published in ref. 39. The authors measured the cloud point temperatures, T_{cloud}, as a function of composition of the mixture. The crystalline proteins γD and βB1 are approximately the same size, what makes our analysis somewhat simpler. Experimental data are converted from the volume packing fraction values ϕ[39] to corresponding η units *via* the relation: $\eta = \pi\phi N_A\sigma^3/[3\Omega(M_1 + x_2(M_2 - M_1))]$, where N_A is Avogadro's number, $M_1 = 20\,607$ g mol^{-1} and $M_2 = 27\,892$ g mol^{-1} molar masses of proteins, and $\Omega = 7.1 \times 10^{-4}$ L g^{-1} the specific volume of proteins.[39]

The best-fit parameters, collected in Table 1, were determined by hand; initially we picked m_1, ε_{11}, and ω_{11} to fit the envelope of pure γD protein ($x_2 = 0.0$). Experimental results for pure βB1 ($x_2 = 1.0$) are not available and to adjust m_2, ε_{22}, and ω_{22} values, we fit the isotherm x_2 *vs.* η at $T = 268.5$ K (see Fig. 3). In this range of x_2 values the T_{cloud} data are readily available.[39] This is all the input information we use. We also assume that conventional

Table 1 Model parameters used to analyze experimental data of Wang[39] *et al.* In agreement with crystallographic data we use σ = 2.95 nm

γD		βB1	
m_1	8	m_2	2
ε_{11}/k_B [K]	2231	ε_{22}/k_B [K]	3248
ω_{11} [nm]	0.18	ω_{22} [nm]	0.21

Lorentz–Berthelot mixing rules provide a reasonable measure of the cross interactions; it is known that an excessive deviations in parameters may cause an un-physical instability of protein mixtures.[59]

Parameters listed in Table 1 are then used to calculate: (i) the isotherm at T = 270.5 K (including the tie-lines), for which the T_{cloud} measurements are available (see Fig. 3); (ii) T_{cloud} *vs.* η curves for x_2 equal to 0.08, 0.16, and 0.29, shown in Fig. 4 as colored curves; (iii) the isotherms at 255 K, 265 K, and 275 K presented in Fig. 5; as also (iv) the equilibrium surface for the model γD–βB1 mixture (see Fig. 4) for a range of temperatures, x_2, and η values.

Results for the calculated tie-lines and isotherms at 268.5 K are, as shown in Fig. 3, in fair agreement with measurements.

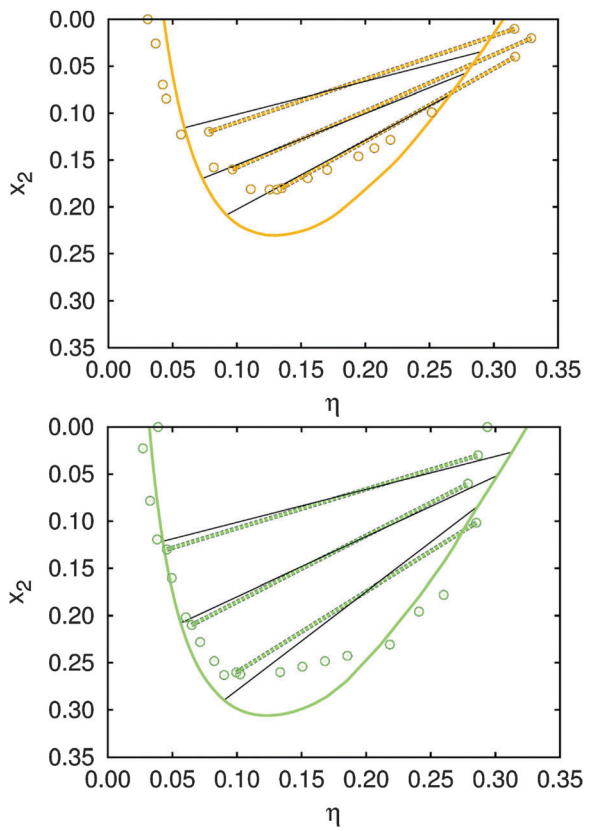

Fig. 3 Isotherms at 268.5 K (green, bottom), and 270.5 K (orange, top). Circles show experimental data, while the solid lines represent calculations. In addition, we compare the experimentally determined tie-lines[39] (dashed green/orange lines), with those calculated (black solid lines). Experimental data for the isotherm at 268.5 K (without the tie-lines) are showed together with the measurements for x_2 = 0 (Fig. 4) used to fix the model parameters, collected in Table 1. Other results, including the isotherm at 270.5 K and the tie-lines for both temperatures, were than calculated using these values.

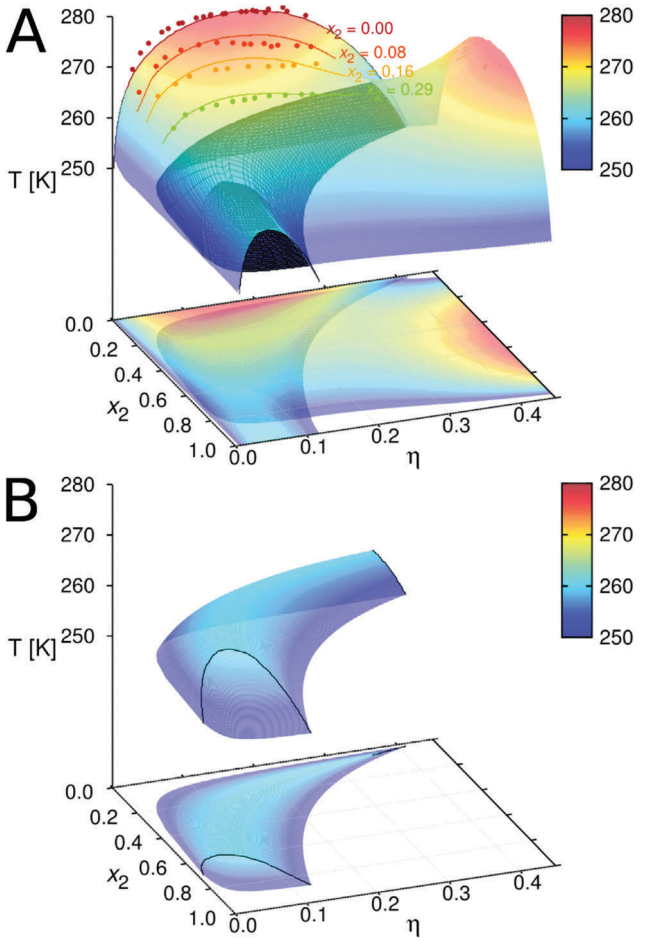

Fig. 4 Panel A: Phase separation of the γD–βB1 mixture in 0.1 M phosphate buffer at pH = 7.0. We show the equilibrium surface and its projections on the η–x_2 plane. Experimental results are denoted by points and calculations by lines. The comparison is shown for x_2 values equal to 0, 0.08, 0.16, and 0.29. The three-phase region (black shaded area) is embedded into the two-phase region for temperatures below 264 K. Panel B: For easier visualization we re-plotted the three-phase region. Coexisting phases are denoted by black lines, see also the isotherm at T = 255 K in Fig. 5.

The fact that we obtained correct slopes for the tie-lines supports the proposed model; it is known that the latter are very sensitive to the protein–protein interactions. As concluded by Dorsaz *et al.*;[6] the difference in cross interactions at less then k_BT may already invert the slope of the tie-line. Notice that uncertainties in measurements of the packing fraction and composition of the coexisting phases can be up to ±0.02.[39]

In Fig. 4 we show the agreement between the T_{cloud} measurements and theoretical calculations (lines) for x_2 values equal to 0.00, 0.08, 0.16, and 0.29. In addition to these results we also show the equilibrium surface, which signals the onset of cloudiness and the coexistence of low- and high-packing fraction (η) liquid phases. A large difference between m_1 (equal to eight), and m_2 (equal to two) for γD and βB1, we picture this situation in Fig. 1, is needed for the model to correctly describe experimental results. This set of m_i values leads to the presence of a three-phase region (black shaded area) at temperatures below 264 K, where no experimental results are available. Notice that the two-phase coexistence region

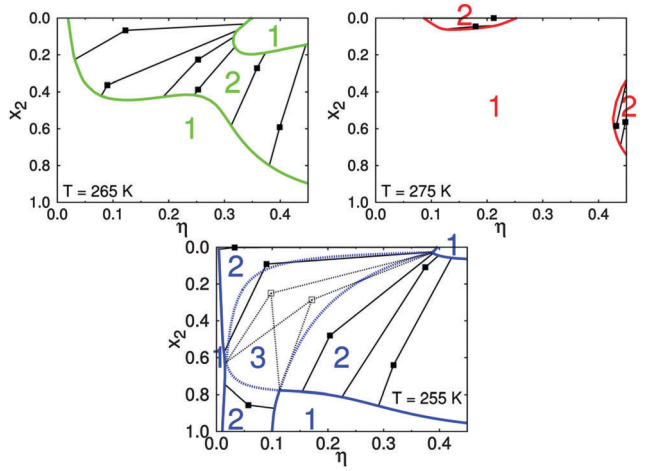

Fig. 5 Isotherms for the temperatures 255 K, 265 K, and 275 K, extracted from Fig. 4. The color notation corresponds to the temperature code in Fig. 4. The one-, two-, and three-phase regions are denoted by the appropriate numbers. Squares denote mixtures, which phase-separate while the lines are connecting the mother phases with the coexisting (daughter) phases.

is present even for packing fractions as high as $\eta > 0.4$, what was not the case when the difference between m_1 and m_2 was smaller (compare Fig. 2). βB1 protein exhibits a slightly longer range ω_{22} and also a somewhat stronger association energy ε_{22}. The measurements of the association enthalpies,[60] based on the thermodynamic analysis of protein oligomerization, reveal a much larger association enthalpy in case of βB1 crystallin. This finding is consistent with our calculations. It is worth mentioning that the potential parameters (ε_{11}, ε_{22} and ω_{11}, ω_{22}) applied in this calculation fall in a broad range of hydrogen bond magnitudes.[61] This result is consistent with our previous findings.[34]

Fig. 5 presents the results for the three isotherms, which are part of the full phase-diagram shown in Fig. 4. We show the η–x_2 projections of the equilibrium surface for three different temperatures: 255 K, 265 K, and 275 K. The graphs show an evolution of the phase coexistence curves with temperature. For the highest one, T = 275 K, the one-phase region is dominant; the two-phase regions only exist around $x_2 = 0$ (η between 0.1 and 0.25) and also for high η values with the composition x_2 between roughly 0.4 and 0.8. For a somewhat lower temperature, T = 265 K, the two-phase region is much broader. The lines indicate how the unstable points (squares) phase separate into the phases having different x_2, η values. For the lowest isotherm, T = 255 K, the graph is quite complex. Here we have a domain, denoted by "3", where three liquid phases coexist. In addition, both one- and two-phase regions are present. The complexity is very likely due to fact that interprotein interactions gain in importance if the temperature is lowered. See also the discussion of results presented in panel C of Fig. 2.

Although the choice of the set of parameters is not unique, there seems to be a limited number of such sets providing reasonable results. In several examples we tested the sensitivity of the parameters listed in Table 1. In one of these cases we assumed for the number of sites to be equal to $m_1 = m_2 = 8$, and ranges of $\omega_{11} = \omega_{22} = 0.18$ nm, while at the same time, we increased the difference between ε_{11} and ε_{22} by 12%. With these modifications, we have not been able to reproduce the experimental isotherm at T = 268.5 K, shown in Fig. 3; the result for $x_2 \approx 0.3$ at $\eta \approx 0.14$ was (in our best fitting attempt) too large by a factor of two. In yet another example we choose $m_1 = 8$ ("fixed" by fitting the experiment at $x_2 = 0$) and $m_2 = 4$, keeping the same temperature and ω_{11} and ω_{22} values as above. Again, even upon considerable adjustment of ε_{22} (we took a 15% more-negative value) the agreement with experimental x_2 *vs.* η curve was poor, much poorer than in Fig. 3.

The authors of the experimental work provided data in the domain $\eta \times x_2 \in [0,0.3] \times [0,0.3]$; the state points out of this range were not accessible to measurements due to the tendency of the system to solidify. We calculated the system properties for η and x_2 values out of this range – no convergence problems were encountered. An example is the liquid–liquid envelope at $x_2 = 1$, *i.e.* for pure βB1 protein. Unfortunately, this curve could not be obtained by direct measurements, because of the low critical temperature below −10 °C.[60] Addition of polyethylene glycol to the protein buffer mixture increases the critical point and allows us to measure the cloud point temperatures in a wide range of protein concentrations. Such indirect measurements,[60] in conjunction with the extrapolation to zero concentration of polyethylene glycol, suggest for the critical point to be around 250 K. Our theoretical model yields for $x_2 = 1$ (see Fig. 4) a value equal to 255 K. Notice again that this prediction is based on fitting the results for $x_2 = 0$ (absence of βB1 protein) and the isotherm for the mixture at T = 268.5 K.

An alternative (and to a certain degree complementary) presentation of the information given by phase diagram in Fig. 4 and isotherms in Fig. 5 is provided by the cloud and shadow curves.[62] These curves are fundamental for studying mixtures, because they show the partitioning of protein species into low- and high-packing fraction phases. These results are shown in the ESI† as Fig. S1.

Relating model parameters to structural information

Crystallographic data for both proteins are available from the Protein Data Bank (PDB) server. While for γD (1*hk*0) complete structure is provided, for βB1 only a truncated structure is available. The truncation applies to the hydrophobic tail (41 amino acids) at the N-terminus, named βB1ΔN41 (1*oki*). Even though the absence of the hydrophobic tail may modify phase behavior,[60] the quasi-elastic light scattering and circular dichroism measurements of the wild-type βB1 and mutant βB1ΔN41 revealed no significant differences.[60]

The surface distributions of charges and insights into tertiary structures of γD and βB1ΔN41 proteins are shown in Fig. 6. Since all the measurements were performed slightly below the isoelectric points of proteins (7.1), at pH equal to 7.0, the net charge of proteins is around zero (blue and red surface areas are roughly of the same size).

From the Fig. 6 it is difficult to draw any firm conclusions about why the numbers of sites per sphere (m_1,m_2), needed to

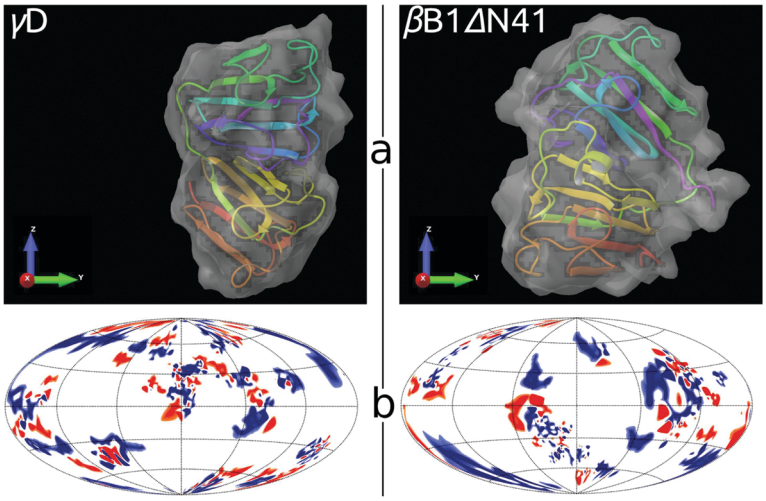

Fig. 6 Comparison between γD (left) and βB1ΔN41 (right) crystallins. Panel a: Tertiary structures embedded into the solvent accessible area, generated by Schrödinger software.[63] Panel b: Charged amino acid residues on the surfaces of proteins; positive – blue, negative – red, neutral – white. The location of residues is shown in the Hammer projection,[64] having the polar and the azimuthal angles as the principle coordinate axes. An implementation of this projection was proposed by Koromyslova *et al.*[65] and modified by us, as described in the ESI.†

Table 2 Monopole (charge), dipole, and quadrupole moments for γD and βB1ΔN41 crystallins from thr Protein Dipole Moments Server, atomic units, see ref. 66

Charge	Dipole	Quadrupole
γD crystalline		
0	191	110
βB1ΔN41 crystallin (chain A)		
0	260	1037

adequately describe crystalline mixture, differ from one protein to another so much. In search for some clues we turned to the Protein Dipole Moments Server[66] (http://dipole.weizmann.ac.il/) for information about higher multipole terms for 1*hk*0 and 1*oki*. The results are summarized in Table 2. The two proteins have dipole moments of similar magnitude, while βB1ΔN41 has a more than nine times larger quadrupole moment than γD.

This enhances the deviation from isotropic behavior and a low number of sites per sphere ($m_2 = 2$) is needed to describe the highly directional interactions among the βB1 proteins. There is experimental evidence showing that β crystallin proteins form dimers and low-branched oligomers.[60,67]

7 Conclusions

We use a simple model to calculate the phase diagram for the mixture of two proteins. Both protein species are modeled as dumbbells (two fused hard spheres) decorated by different numbers of attractive square-well sites allowing inter-protein attraction. The major parameters of the model are the number of attractive sites per sphere forming the protein (m_i; $i = 1, 2$) and their attractive range and square-well depth. In the first part of the study we investigated how the principal model parameters shape the phase diagram. This initial calculations showed that the number of attractive sites per sphere of protein plays an important role. In the second part of the study we turned to analyze the behavior of the lens proteins. The mixture of γD and βB1 crystallins was thoroughly studied experimentally by Wang *et al.*,[39] and the authors provided the cloud point temperatures for various experimental conditions. To fix the principal parameters of our model – in particular m_1 and m_2 and the attractive square-well depths – we fitted the experimental results for pure γD protein and the isotherm for the mixture at $T = 268.5$ K. The cross interaction between proteins was described by the usual Lorentz–Berthelot rule. Numerical evaluation of this model was performed by Wertheim's thermodynamic perturbation theory, which allows calculation of the measurable properties, including the complete phase diagram.

The comparison of the model calculations with experimental data for binary mixture of γD and βB1 crystallins yields the following conclusions: (i) for good agreement between theory and experiment a large difference in number of sites per sphere of protein ($m_1 = 8$ and $m_2 = 2$) is needed. The results are consistent with experimental findings indicating that β crystallins form small clusters with only a few proteins involved. We trace this result to the large difference in the quadrupole moments between the two proteins. (ii) Our model predicts correct slopes for the tie-lines in x_2 *vs.* η graphs. (iii) The calculation yields good agreement with experimentally determined temperature-packing fraction curves for mixtures. (iv) The model provides the results for pure protein βB1, for which no direct measurements are available, and accordingly these data were not included in the fitting procedure. The calculation for the critical temperature is in good agreement with the value suggested from the extrapolation of experimental data for the βB1 protein. (v) The "dumbbell" model predicts a region where three liquid phases are in equilibrium – the region is just out of the domain where measurements are currently available. The calculation demonstrates, in agreement with articles listed in recent review papers,[7,8] that methods of the condensed matter physics may be useful in the analysis of experimental data of protein systems.

Acknowledgements

This study was supported by the Slovenian Research Agency fund (ARRS) through Program 0103-0201, by the NIH research grant (GM063592), and the Young Researchers Program (M. K.) of the Republic of Slovenia.

References

1 *Protein Condensation: Kinetic Pathways to Crystallization and Disease*, ed. J. D. Gunton, A. Shiryayev and D. L. Pagan, Cambridge University Press, 2007.
2 S. E. Ahnert, J. A. Marsh, H. Hernández, C. V. Robinson and S. A. Teichmann, *Science*, 2015, **350**, 2245.
3 G. M. Tavares, T. Groguennes, P. Hamon, A. F. Carvalho and S. Bouhallab, *Food Hydrocolloids*, 2015, **48**, 238–247.
4 A. Kurut, B. A. Persson, T. Åkersson, J. Forsman and M. Lund, *J. Phys. Chem. Lett.*, 2012, **3**, 731–734.
5 N. Dorsaz, G. M. Thurston, A. Stradner, P. Schurtenberger and G. Foffi, *J. Phys. Chem. B*, 2009, **113**, 1693–1709.
6 N. Dorsaz, G. M. Thurston, A. Stradner, P. Schurtenberger and G. Foffi, *Soft Matter*, 2011, **7**, 1763–1776.
7 D. Fusco and P. Charbonneau, *Colloids Surf., B*, 2016, **137**, 22–31.
8 J. J. McManus, P. Charbonneau, E. Zaccarelli and N. Asherie, *Curr. Opin. Colloid Interface Sci.*, 2016, **22**, 73–79.
9 A. L. Fink, *Folding Des.*, 1998, **3**, R9–R23.
10 W. Wang, *Int. J. Pharm.*, 2005, **289**, 1–30.
11 S. Frokjaer and D. E. Otzen, *Nat. Rev.*, 2005, **4**, 298–306.
12 A. George and W. W. Wilson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1994, **50**, 361–365.
13 G. A. Vliegenthart and H. N. W. Lekkerkerker, *J. Chem. Phys.*, 2000, **112**, 5364–5369.
14 G. B. Benedek, *Invest. Ophthalmol. Visual Sci.*, 1997, **38**, 1911–1921.
15 C. A. Ross and M. A. Poirier, *Nat. Med.*, 2004, **10**, S10–S17.
16 A. M. Thayer, *Chem. Eng. News*, 2016, **94**, 30–35.
17 *Theory of the Stability of Lyophobic Colloids*, ed. E. J. W. Verwey and J. T. G. Overbeek, Elsevier, 1948.
18 R. P. Sear, *J. Chem. Phys.*, 1999, **111**, 4800–4806.
19 V. J. Anderson and H. N. W. Lekkerkerker, *Nature*, 2002, **416**, 811–815.
20 E. Bianchi, R. Blaak and C. N. Likos, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6397–6410.
21 M. Lund, *Colloids Surf., B*, 2016, **137**, 17–21.
22 H. Liu, S. K. Kumar and F. Sciortino, *J. Chem. Phys.*, 2007, **127**, 084902.
23 C. Gögelein, G. Nägele, R. Tuinier, T. Gibaud, A. Stradner and P. Schurtenberger, *J. Chem. Phys.*, 2008, **129**, 085102.
24 A. Fortini, E. Sanz and M. Dijkstra, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2008, **78**, 041402.
25 D. Fusco and P. Charbonneau, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2013, **88**, 012721.
26 Y. U. Moon, C. O. Anderson, H. W. Blanch and J. M. Prausnitz, *Fluid Phase Equilib.*, 2000, **168**, 229–239.
27 J. M. Tavares, P. I. C. Teixeira, M. M. T. da Gama and F. Sciortino, *J. Chem. Phys.*, 2010, **132**, 234502.
28 D. J. Audus, F. W. Starr and J. F. Douglas, *J. Chem. Phys.*, 2016, **144**, 074901.
29 D. de las Heras, J. M. Tavares and M. M. T. da Gama, *J. Chem. Phys.*, 2011, **134**, 104904.
30 D. de las Heras, J. M. Tavares and M. M. T. da Gama, *Soft Matter*, 2011, **7**, 5615–5626.
31 S. Roldnán-Vargas, F. Smallenburg, W. Kob and F. Sciortino, *J. Chem. Phys.*, 2013, **139**, 244910.
32 M. Kastelic, Y. V. Kalyuzhnyi and V. Vlachy, *Condens. Matter Phys.*, 2016, **19**, 23801.
33 Y. Wang, A. Lomakin, R. F. Latypov and G. B. Benedek, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 16606–16611.
34 M. Kastelic, Y. V. Kalyuzhnyi, B. Hribar-Lee, K. A. Dill and V. Vlachy, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 6766–6770.
35 T. Janc, M. Kastelic, M. Bončina and V. Vlachy, *Condens. Matter Phys.*, 2016, **19**, 23601.
36 M. S. Wertheim, *J. Stat. Phys.*, 1986, **42**, 477–492.
37 P. S. Sarangapani, S. D. Hudson, R. L. Jones, J. F. Douglas and J. A. Pathak, *Biophys. J.*, 2015, **108**, 724–737.
38 J. M. Prausnitz, *Biophys. J.*, 2015, **108**, 453–454.
39 Y. Wang, A. Lomakin, J. J. McManus, O. Ogun and G. B. Benedek, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 13282–13287.
40 E. Serebryany and J. A. King, *Prog. Biophys. Mol. Biol.*, 2014, **115**, 32–41.
41 M. L. Broide, C. R. Berland, J. Pande, O. O. Ogun and G. B. Benedek, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 5660–5664.
42 P. R. Banerjee, A. Pande, J. Patrosz, G. M. Thurston and J. Pande, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 574–579.
43 S. James, M. K. Quinn and J. J. McManus, *Phys. Chem. Chem. Phys.*, 2015, **17**, 5413.
44 M. K. Quinn, N. Gnan, S. James, A. Ninarello, F. Sciortino, E. Zaccarelli and J. J. McManus, *Phys. Chem. Chem. Phys.*, 2015, **17**, 31177.
45 A. Basak, O. Bateman, C. Slingsby, A. Pande, N. Asherie, O. Ogun, G. B. Benedek and J. Pande, *J. Mol. Biol.*, 2003, **328**, 1137–1147.
46 A. Pande, J. Pande, N. Asherie, A. Lomakin, O. Ogun, J. King and G. B. Benedek, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 6116–6120.
47 P. Evans, K. Wyatt, G. J. Wistow, O. A. Bateman, B. A. Wallace and C. Slingsby, *J. Mol. Biol.*, 2004, **343**, 435–444.
48 J. J. McManus, A. Lomakin, O. Ogun, A. Pande, M. Basan, J. Pande and G. B. Benedek, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 16856–16861.
49 F. Ji, J. Jung and A. M. Gronenborn, *Biochemistry*, 2012, **51**, 2588–2596.
50 *Theory of Simple Liquids*, ed. J. P. Hansen and I. R. McDonald, Elsevier, 2006.
51 J. L. Lebowitz, *Phys. Rev.*, 1964, **133**, A895–A899.
52 G. Jackson, W. G. Chapman and K. E. Gubbins, *Mol. Phys.*, 1988, **65**, 1–31.
53 W. G. Chapman, G. Jackson and K. E. Gubbins, *Mol. Phys.*, 1988, **65**, 1057–1079.
54 E. Bianchi, P. Tartaglia, E. Zaccarelli and F. Scortino, *Chem. Rev.*, 2008, **128**, 144504.

55 M. Emelianenko, Z. K. Liu and Q. Du, *Comput. Mater. Sci.*, 2006, **35**, 61–74.
56 *Numerical Recipes in Fortran 77*, ed. W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, Cambridge University Press, 1992.
57 E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli and F. Sciortino, *Phys. Rev. Lett.*, 2006, **97**, 168301.
58 H. Liu, S. K. Kumar, F. Sciortino and G. T. Evans, *J. Chem. Phys.*, 2009, **130**, 044902.
59 A. Stradner, G. Foffi, N. Dorsaz, G. Thurston and P. Schurtenberger, *Phys. Rev. Lett.*, 2007, **99**, 198103.
60 O. Annunziata, A. Pande, J. Pande, O. Ogun, N. H. Lubsen and G. B. Benedek, *Biochemistry*, 2005, **44**, 1316–1328.
61 D. Xu, C. J. Tsai and R. Nussinov, *Protein Eng., Des. Sel.*, 1997, **10**, 999–1012.
62 P. Sollich, *J. Phys.: Condens. Matter*, 2002, **14**, R79–R117.
63 *Schrödinger Release 2015-2, Maestro, version 10.2*, Schrödinger, LLC, New York, NY, 2015.
64 *Map Projections: A Working Manual*, ed. J. P. Snyder, U.S. Geological Survey, Washington, DC, 1987.
65 A. D. Koromyslova, A. O. Chugunov and R. G. Efremov, *J. Chem. Inf. Model.*, 2014, **54**, 1189–1199.
66 C. E. Felder, J. Prilusky, I. Silman and J. L. Sussman, *Nucleic Acids Res.*, 2007, **35**, W512–W521.
67 O. A. Bateman, N. H. Lubsen and C. Slingsby, *Exp. Eye Res.*, 2001, **73**, 321–331.