Assessing the Intricate Balance of Intermolecular Interactions upon Self-Association of Intrinsically Disordered Proteins

Ellen Rieloff [1], **Mark D. Tully** [2] **and Marie Skepö** [1]

*1 - **Theoretical Chemistry,** Lund University, POB 124, SE-221 00 Lund, Sweden*
*2 - **European Synchrotron Radiation Facility (ESRF),** Grenoble, France*

Correspondence to Ellen Rieloff and Marie Skepö: *ellen.rieloff@teokem.lu.se, marie.skepo@teokem.lu.se*
https://doi.org/10.1016/j.jmb.2018.11.027
Edited by Monika Fuxreiter

Abstract

Attractive interactions between intrinsically disordered proteins can be crucial for the functionality or, on the contrary, lead to the formation of harmful aggregates. For obtaining a molecular understanding of intrinsically disordered proteins and their interactions, computer simulations have proven to be a valuable complement to experiments. In this study, we present a coarse-grained model and its applications to a system dominated by attractive interactions, namely, the self-association of the saliva protein Statherin. SAXS experiments show that Statherin self-associates with increased protein concentration, and that both an increased temperature and a lower ionic strength decrease the size of the formed complexes. The model captures the observed trends and provides insight into the size distribution. Hydrophobic interaction is considered to be the major driving force of the self-association, while electrostatic repulsion represses the growth. In addition, the model suggests that the decrease of association number with increased temperature is of entropic origin.

Introduction

Intrinsically disordered proteins (IDPs) are characterized by a lack of stable tertiary structure under physiological conditions *in vitro* [1,2] and hence are best described by conformational ensembles [3,4]. Bioinformatic studies have led to the conclusion that 10%–20% of the eukaryotic proteins are intrinsically disordered, and even more proteins contain intrinsically disordered regions (IDRs) [5–8]. It has also been established that IDPs and IDRs are involved in many biological processes and diseases, and that the lack of folded structure is related to their functions [7,9].

Attractive interactions between IDPs can lead to the formation of aggregates, which in the case of diseases such as Parkinson's disease and Alzheimer's disease is harmful [10]. IDP attractions can also be fundamental for a desired outcome, such as in the formation of proteinaceous membrane-less organelles [11–14], which are condensed liquid droplets often enriched in IDPs and IDRs and commonly found in the cell cytoplasm and nucleus [15]. Various pieces of evidence suggest that liquid–liquid phase separation is a driving force for the formation of some proteinaceous membrane-less organelles [11–14], and that the phase separation itself is driven by weak multivalent interactions between disordered proteins [16,17].

For understanding IDPs and their interactions, computer simulations are a useful complement to experiments [18,19]. There have been considerable advances regarding atomistic simulations of IDPs, where development and justification of force fields and water models have been validated against experimental results [20–24]. The full-atom approach and explicit water treatment in atomistic simulations are great advantages for gaining a molecular understanding, however, atomistic simulations are computationally demanding, both regarding execution time and data storage. Hence, this poses limitations on the accessible timescale and system size, and therefore, a coarse-grained approach is a more viable option for studying complex systems, such as the examples above. Recently, a coarse-grained model based on the primitive model, in combination with Monte Carlo simulations, has proven capable of capturing bulk properties at dilute conditions for a range of IDPs [25]. We aim to develop this model to also account for more complex systems, and first is the investigation of a model system dominated by intermolecular attractions, namely, the self-association of the saliva protein Statherin. Statherin has a distinct amphiphilic character in its primary sequence, shown in Fig. 1. Almost all charges are located in the N-terminal part, starting with a block of negative charges, followed by a block of positive charges. From the hydropathy values in the Kyte–Doolittle scale [26], it is shown that overall the hydropathy is rather low, which is typical for IDPs. However, residues 15–43 contain seven tyrosines, whose aromatic side chains have been established to be of importance for liquid–liquid phase separation [27,28]. Statherin also consists of 16% proline residues, which are denoted as “disorder-promoting” [29].

In this work, Statherin is characterized experimentally at monomeric conditions through the use of small-angle X-ray scattering (SAXS) and circular dichroism (CD), and at self-associating conditions through SAXS experiments and simulations. The simulation model is validated against the experiments and is demonstrated to be useful for describing polydispersity and the interplay between electrostatics, hydrophobic interactions, and entropy in the self-association process.

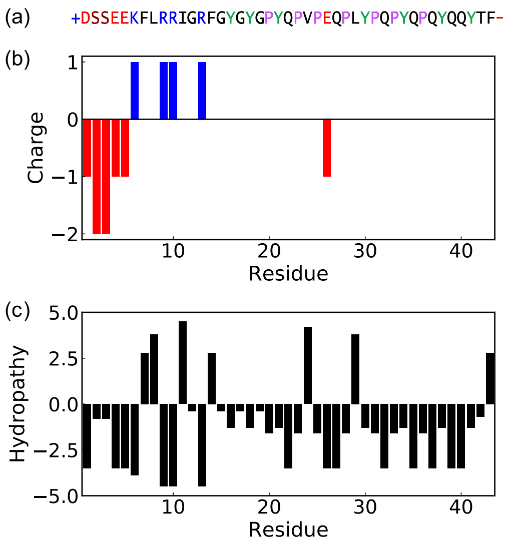

Fig. 1. (a) Amino acid sequence of Statherin with the charge distribution at pH 8 and certain amino acids highlighted. Positive residues are marked in blue, negative in red, phosphorylated serines with the charge −2*e* in dark red, and prolines in lilac and tyrosines in green. (b) Charge distribution and (c) hydropathy values using the Kyte–Doolittle scale, where −4.5 is the most hydrophilic and +4.5 is the most hydrophobic [26].

Results and Discussion

The experimental results for Statherin at monomeric conditions are presented first, followed by the self-association studied both experimentally and by Monte Carlo simulations.

Monomeric behavior

In Fig. 2a–c, data for monomeric Statherin obtained by SAXS coupled with size-exclusion chromatography (SEC' taken from Ref. [25]) is presented. From regular SAXS measurements at low protein concentration (0.24 mg/mL), the molecular weight was determined to be 5.29 kDa, based on the forward scattering, I_0, obtained from the pair distance distribution function, $P(r)$ [25]. This is in good agreement with the theoretical molecular weight of 5.38 kDa, confirming monomeric conditions. As seen in Fig. 2a, Statherin shows the typical featureless scattering profile of an IDP. The IDP character is also verified by the dimensionless Kratky plot in Fig. 2b, where the profile has an uprise slope and reaches a plateau at higher q values, typical for flexible chains. In addition, the CD data presented in Fig. 2d confirm a random coil behavior with some presence of secondary structure. The global minimum is located at 205 nm, which is slightly higher than the usual 198 nm for random coils; however, it is typical for poly-proline II (PPII) structure. The shallow minimum close to 222 nm might suggest a small presence of α-helix. Several studies of Statherin with CD or NMR have suggested that the charged N-terminal has a propensity for forming α-helix and that a part of the middle adopt PPII structure. Nevertheless, the overall structure is still disordered in aqueous solution [30–34]. Fig. 2d also shows that there are no large differences in structure due to salt concentration.

The radius of gyration for monomeric Statherin in 150 mM NaCl has been reported as 19.3 ± 0.2 Å, based on the $P(r)$ presented in Fig. 2c [25]. With urea, the radius of gyration is increased to 22.1 ± 0.2 Å for 4 M urea and to 23.7 ± 0.3 Å for 8 M urea. The dimensionless Kratky plot, shown in Fig. 3a, also indicates an increase in stiffness when urea is added. From CD measurements it is seen that the mean residue ellipticity ($[\theta]_{\mathrm{MRW}}$) at 228 nm, presented in Fig. 3b and c, increases linearly with increased urea concentration and also becomes positive at high urea concentrations. This corresponds to an increase of PPII content, in agreement with the study by Whittington *et al.* [35], reporting that urea promotes PPII formation. PPII conformation is more extended than both random coil and α-helix; hence, this explains the changes observed in the SAXS measurements.

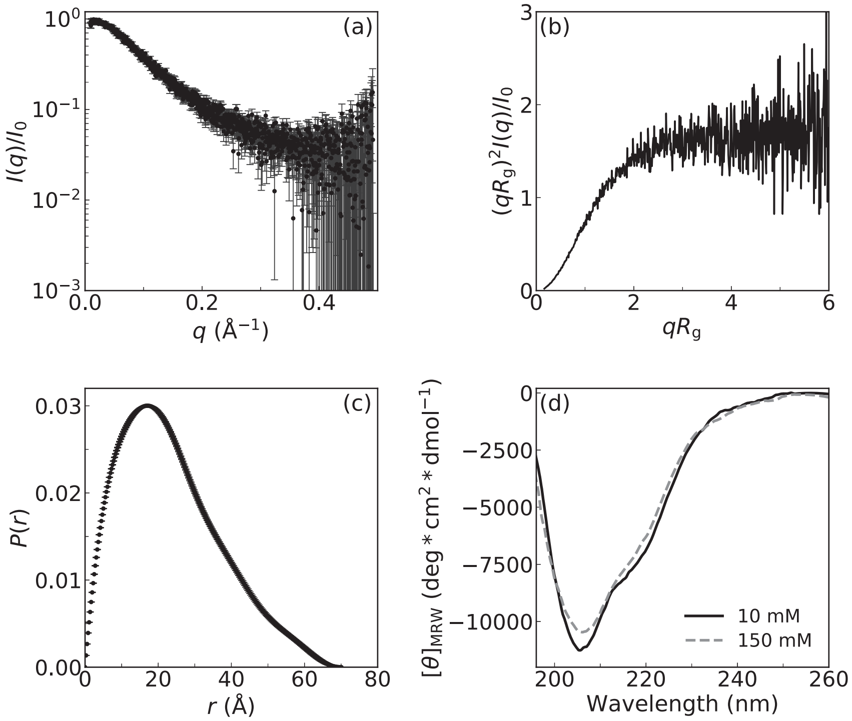

Fig. 2. SAXS data for Statherin obtained by SEC-SAXS, at 150 mM NaCl and 20 mM Tris buffer with pH 8, from Ref. [25]. (a) Form factor, (b) dimensionless Kratky plot, and (c) pair distance distribution function. (d) CD spectra for Statherin in 10 and 150 mM NaF and 20 mM phosphate buffer (pH 8) with a protein concentration of 0.11 and 0.13 mg/mL, respectively, measured at 20 °C.

Temperature also induces changes in secondary structure. With increased temperature, the $[\theta]_{MRW}$ increases at 205 nm and decreases at 228 nm, as shown in Fig. 4, suggesting a loss of PPII as described by Kjaergaard *et al.* [36] for other IDPs. The loss of PPII appears rather proportional to temperature.

Self-association

Experimental results

With increased protein concentration, Statherin self-associates into complexes, which is evident from an increase in forward scattering. The average number of proteins per complex was determined from the forward scattering and is presented against the protein concentration in Fig. 5a for the reference system with 150 mM NaCl. Panels b and d in the same figure present corresponding data from simulations and will be discussed in the next section. The growth is linear with respect to concentration up to 10 mg/mL, and afterward, the slope decreases, which might suggest a maximum size of the Statherin complex. Likewise, the radius of gyration follows the same trend, although a plateau is reached earlier. However, a depression of the forward scattering at higher concentrations due to a structure factor cannot be ruled out, and therefore, the high concentration data should be interpreted with care. Especially since, at 24 mg/mL and higher concentrations, inter-particle interference is visible in the *P*(*r*) as a decrease below zero at long distances. The scattering curves, Guinier plots, and I_0 and radius of gyration determined by both Guinier and *P*(*r*) are provided in Supplemental information.

The Kratky plot in Fig. 5c shows a transition from flexible chain behavior to more globular when the complexes are formed. The complexes are also more spherical in shape than the free proteins, which is evident from the pair distance distribution function presented in Fig. 6, plotted to enhance the differences compared to a sphere.

Since urea weakens hydrophobic interactions [37], the effect of urea on the Statherin complexes was studied. With 8 M urea, no increase in forward scattering was observed even when reaching 32 mg/mL in protein concentration. The only effect observed was a lowering of the forward scattering due to a structure factor emerging. This indeed suggests hydrophobic interactions as a driving force for the self-association in Statherin. With 4 M urea, it was a downshift at intermediate *q* when going from 2 to 4 mg/mL and that continued for even higher protein concentrations (data not shown). This in

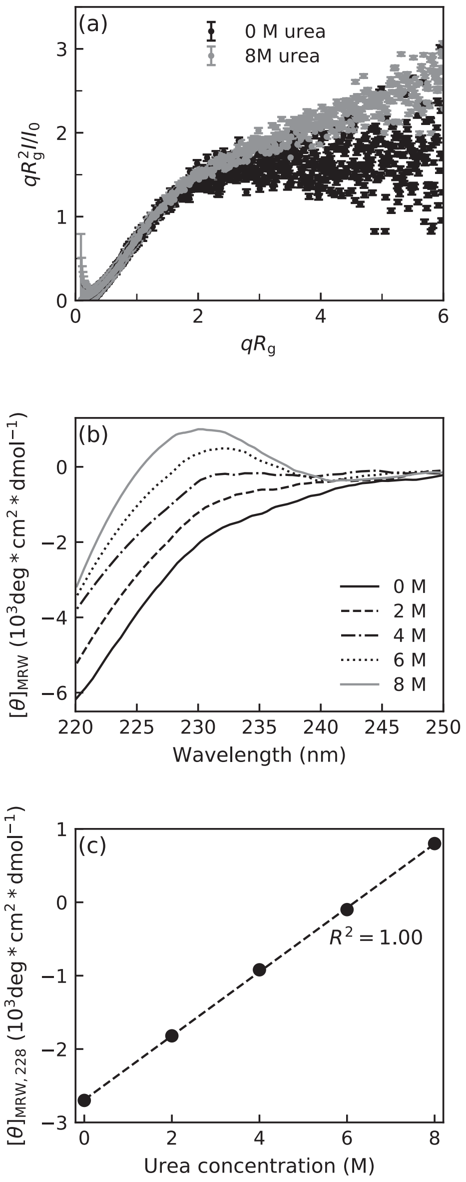

Fig. 3. Effect of urea. (a) Dimensionless Kratky plot for Statherin at 150 mM NaCl measured by SEC-SAXS and with 8 M urea measured by SAXS at a protein concentration of 4 mg/mL, (b) CD spectra and (c) mean residue ellipticity at 228 nm for Statherin (0.12–0.14 mg/mL) *versus* urea concentration, obtained from CD measurements at 20 °C and pH 8.

combination with a decrease in slope in the Kratky plot with increasing concentration suggests that there are still complexes forming in 4 M urea. For surfactants, both the critical micelle concentration and the micelle size have been reported to change with the concentration of urea [38–40].

Self-association has been observed no matter the salt concentration, which supports hydrophobic interactions being the major driving force. However, the average association number appears to increase with increased ionic strength, as presented in Fig. 7a. Due to the possibility of structure factor influence on the scattering data at lower ionic strength, the effect of electrostatic interactions is further discussed within the framework of the simulations (data presented in Fig. 7b).

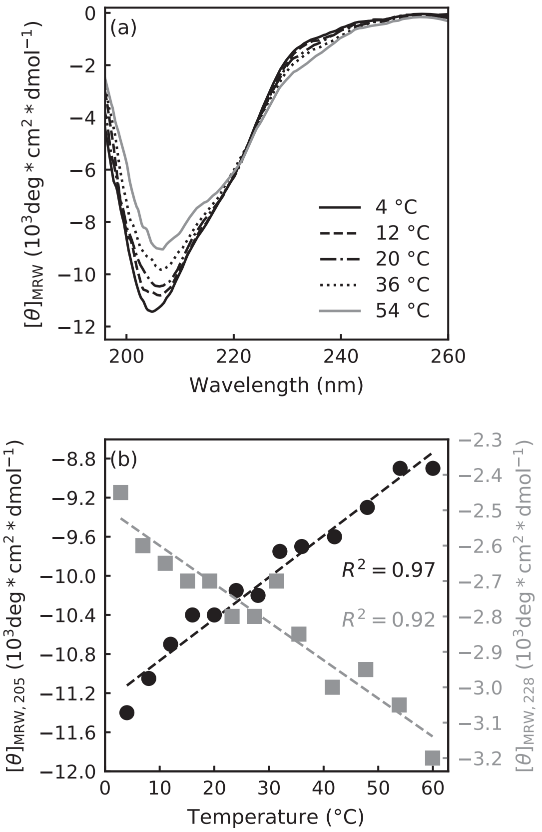

Fig. 4. Temperature dependence of monomeric Statherin (0.13 mg/mL) with 150 mM NaF in 20 mM phosphate buffer at pH 8. (a) CD spectra and (b) mean residue ellipticity at 205 nm (black circles) and 228 nm (gray squares).

Changing the temperature also affects the self-association, as shown by a decrease in association number with increased temperature in Fig. 8. The average radius of gyration follows the same trend (data not shown). The decrease of the association number with temperature has also been observed for surfactants with ionic or zwitterionic headgroups [41], while non-ionic surfactants have shown the opposite temperature dependence [41,42]. For the intrinsically disordered milk-protein β-casein, the association number increases with increased temperature at neutral pH [43], as for non-ionic surfactants. Although β-casein and Statherin have similar block structures, the overall hydrophobicity is higher in β-casein. Hence, it is not unreasonable that the temperature dependence is different.

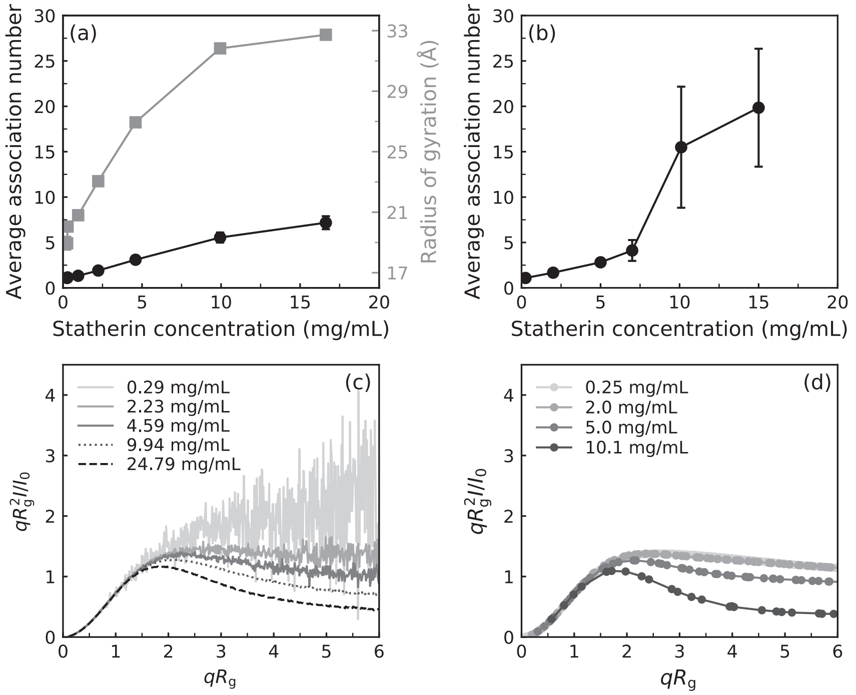

Fig. 5. (a) Average number of proteins per complex (black circles) and radius of gyration (gray squares) *versus* protein concentration determined from SAXS. (b) Average number of proteins per complex *versus* protein concentration from simulations. (c) Dimensionless Kratky plot from experiments. (d) Dimensionless Kratky plot from simulations. The data is reported for the reference system (experimental conditions: 20 mM Tris, 150 mM NaCl, pH 8, 20 °C; simulation conditions: 150 mM implicit salt, 20 °C). In panel a, the error bars on the association number represent a 10% uncertainty.

Simulation results

We have simulated the Statherin system using a modified version of the coarse-grained model presented in Ref. [25]. Therein it was shown that the coarse-grained model works well for Statherin at monomeric conditions. However, to capture the self-association, an additional attractive interaction is needed. We have implemented a short-ranged potential corresponding to 1.32 *kT* at closest contact between neutral amino acids, mimicking a smeared hydrophobic interaction, which causes the proteins to associate upon increased concentration. For the reference system, 150 mM salt, the simulation data follow the linear trend described in experimental data up to approximately 7 mg/mL, according to Fig. 5b. Then it deviates, by forming large complexes, which shall be interpreted as that the model is reliable only at lower protein concentrations. The model is able to capture the experimentally established transition to a more globular state with increased protein concentration in the Kratky plot, c.f. Fig. 5d and c, although the single chain is too compact due to the extra attraction. To capture the behavior at both monomeric conditions and higher protein concentrations, an angular potential can be included as well. However, since the goal with this model is to capture general trends, an exact matching with the experimental Statherin data is not important, and hence, the results of the model without further modifications are presented.

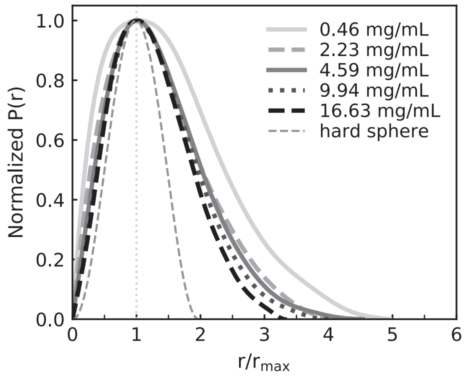

Fig. 6. Pair distance distribution function normalized to enhance deviations in shape from a homogeneous hard sphere, where r_{max} corresponds to the value of r where $P(r)$ has its maximum, for the reference system (20 mM Tris, 150 mM NaCl, pH 8, 20 °C).

The simulations show that the complexes are polydisperse; see the complex size probability distribution in Fig. 9a. At 7 mg/mL and lower concentrations,

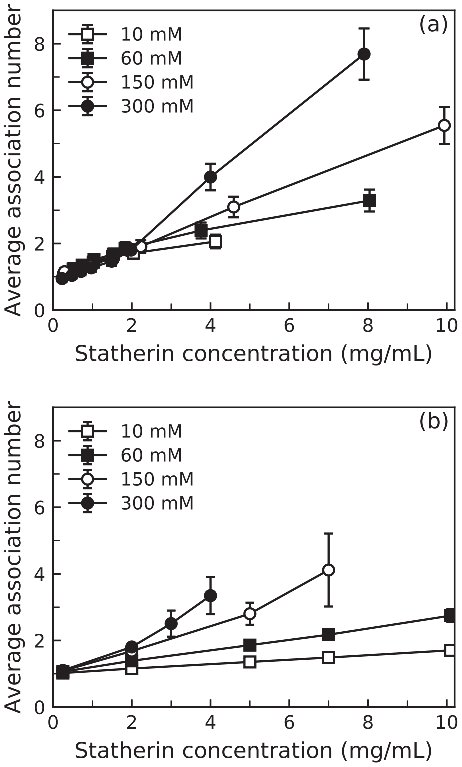

Fig. 7. Average association number determined (a) by SAXS and (b) from simulations, as a function of Statherin concentration for different concentrations of NaCl, at 20 °C. The error bars in panel a represent a 10% uncertainty.

the monomer is the dominating specie and the amount of the different species decreases with increasing size. The polydispersity and monomeric dominance is also evident from the snapshot in Fig. 9b, which furthermore suggests that it is the middle and C-terminal part that forms the core of the complex and that the charged N-terminal part is located on the surface of the complex.

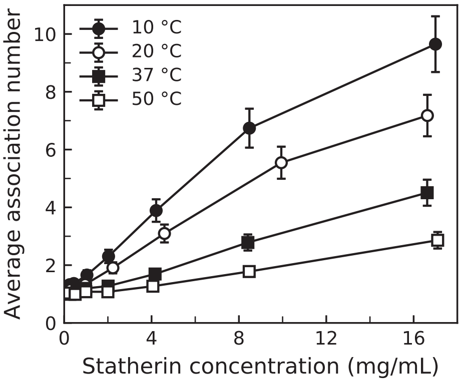

Fig. 8. Average number of proteins per complex determined by SAXS *versus* protein concentration at 150 mM NaCl for 10–50 °C. The error bars represent a 10% uncertainty. The data at 20 °C correspond to the data at 150 mM NaCl in Fig. 7a.

The contact probability between residues of different chains is presented in Fig. 9c and confirms indeed that it is the neutral amino acids that are mostly in contact with other chains. In Fig. 9d, the radial number density distribution from the complex center of mass is presented. It again confirms that the core consists of neutral residues. The negatively charged residue 26 is also part of the core of the complex. The other charged residues are located closer to the surface of the complex.

The experimental $P(r)$ in Fig. 5d shows that the complexes are more spherical than the monomers, due to the change with increasing concentration. However, the experiments only provide the average over all different complex sizes. In the simulations, we have calculated the principal moments of the gyration tensor and from that the asphericity for the complexes of different sizes. It indeed confirms that the monomers are not spherical, having an asphericity value of 0.41. The asphericity decreases with increasing association number until six, where it stabilizes around 0.13 also for larger complexes. If the asphericity is less than 0.1, the object is normally considered spherical [44]. The decrease in asphericity agrees with the experimental results and furthermore shows that the complexes are close to the spherical limit. However, for complexes consisting of seven protein chains, $\langle R_1^2 \rangle$, $\langle R_2^2 \rangle$ and $\langle R_3^2 \rangle$ were 323.5 ± 7.1 Å^2, 158.2 ± 1.2 Å^2, and 91.1 ± 0.5 Å^2, respectively, showing that the instantaneous shapes of the complexes are still not spherical.

The increase of size of the complexes with increased ionic strength observed in SAXS experiments is also captured by the simulations, as seen in Fig. 7b, even if the effect is slightly overestimated compared to experiments (Fig. 7a). This confirms that although the hydrophobic interaction is the major driving force for self-association, electrostatic repulsion stabilizes the system and depresses the growth. To further investigate the electrostatic effect, we performed simulations without phosphorylated serines, which increases the net charge from −4 to 0. This shifts the complex size probability distribution toward larger sizes, depicted in Fig. 10. The overall contact probability also increases from 0.36 ± 0.03 with phosphorylated serines to 0.41 ± 0.01 without phosphorylations at a protein concentration of 2 mg/mL, while the contact profile remains similar in shape. This demonstrates that phosphorylations indeed affect the electrostatic interactions and that it is of importance for the self-association.

Another mutation that illustrates the importance of electrostatics is the point mutation of residue 26, glutamic acid, changing the negatively charged residue located in the middle of the neutral block to a neutral residue. Already in a simulation at 2 mg/mL, the majority of the chains join in one large complex, while for comparison, the reference system rarely exhibits complexes larger than tetramers at the same

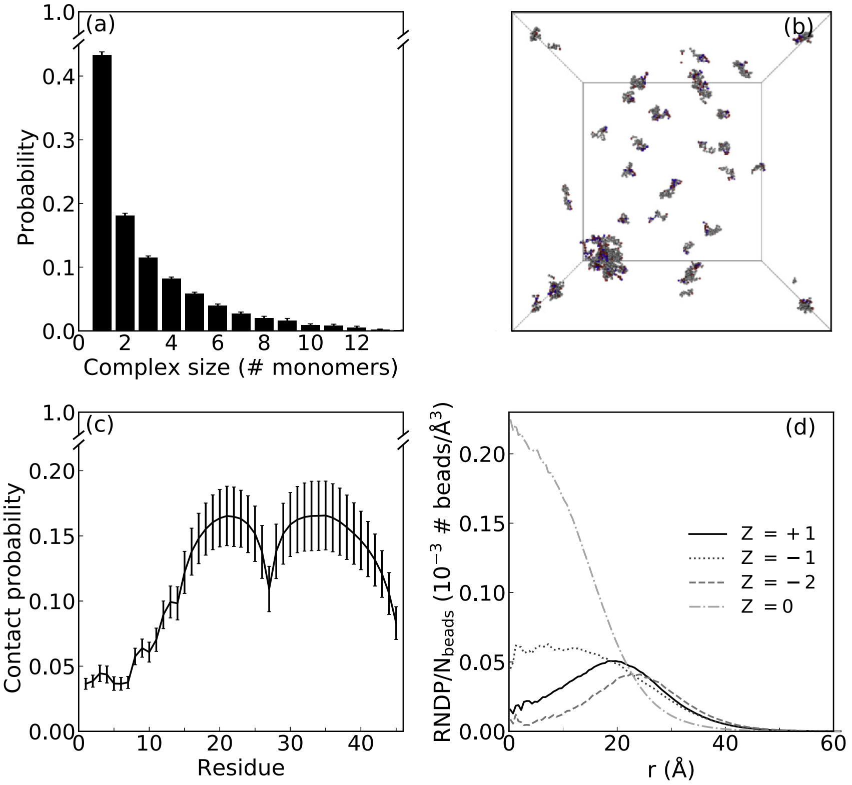

Fig. 9. Simulation data at 5 mg/mL with 150 mM implicit salt. (a) Complex size probability distribution. (b) Snapshot with excluded counterions, where gray beads represent neutral residues, red beads represent negatively charged residues, and blue beads represent positively charged residues. (c) Chain contact probability profile. (d) Radial number density for different bead types, normalized by the number of beads of each type in the protein, as a function of distance from the core center of mass, for complexes consisting of seven proteins. Z represents the charge of each bead type.

concentration. This shows that specific residues can make a great difference for the self-association (results not shown).

With increased temperature, the average association number, displayed in Fig. 11, decreases, again in accordance with experimental results. Since Statherin has a net charge of $-4e$, the overall electrostatic interaction is repulsive. Increased temperature enhances electrostatic interactions, and hence, it would counteract self-association by enhancing the net electrostatic repulsion between Statherin monomers. In addition, the effect of entropy, also opposing self-association, increases with temperature as well. Note that the hydrophobic interaction is regarded temperature-independent in this model. Simulations of the Statherin system without charges at a concentration of 4 mg/mL show a decrease in average association number between 20 and 50 °C, from 3.06 ± 0.63 to 1.39 ± 0.01, compared to 2.24 ± 0.15 to 1.40 ± 0.01 for the same system with charges. This suggests entropy as the main contribution to the temperature effect.

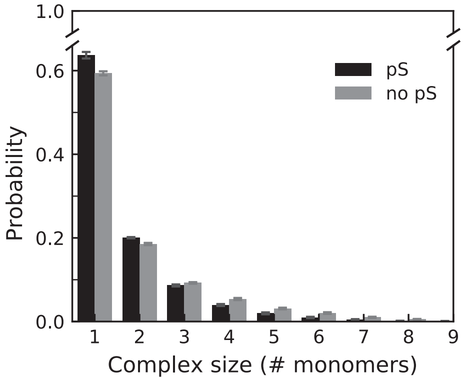

Fig. 10. Complex size probability distribution for 2 mg/mL Statherin with and without phosphorylated serines at 150 mM ionic strength.

Temperature also affects the structure of the complexes. Overall, the asphericity increases as a function of temperature for complexes of the same size, as seen in Fig. 11b. In addition, the radius of gyration also shows the same trend, for example, for complexes of seven proteins, the R_g goes from 22.8 ± 0.1 to 29.8 ± 0.2 Å when temperature changes from 15 to 50 °C. These changes reflect an

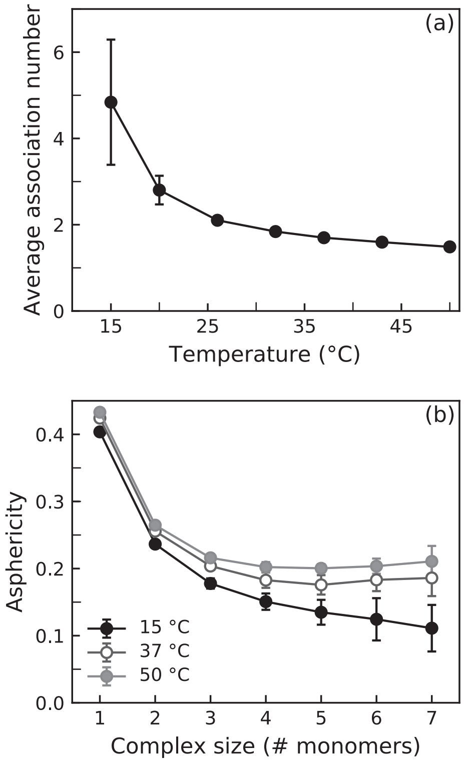

Fig. 11. (a) Average association number as a function of temperature at 5 mg/mL. (b) Asphericity *versus* association number at 15, 37 and 50 °C.

increased flexibility in the complexes, which is expected due to the entropy increase. Although it was shown in the monomeric section that the structure of the individual protein chain changes upon temperature increase, it is expected to be of minor importance for the self-association process, due to the model capturing the trends without including such detail.

Model limitations and improvements

From the simulations, it is apparent that the model breaks down at higher concentrations. The exact concentration depends on the conditions, especially temperature and ionic strength. At the lower-salt concentrations (10 and 60 mM), no breakdown is observed even at 20 mg/mL. The breakdown can be connected with the implicit treatment of salt, since simulations with 150 mM explicit salt and 20 mg/mL protein or more still give an average size less than 10 chains/complex. Hence, an explicit treatment of electrostatics is suggested to provide better results, although at a high computational cost. In the model, the hydrophobic interaction, mimicking the effect of both the enthalpic contribution and the entropic effect on the water molecules, is regarded temperature independent. Including temperature dependence would change the exact values to a certain extent, although the trend would remain. Hence, it would not affect the conclusion that entropy in the system is the largest contributor to the temperature effect for this protein.

Conclusions

A modified version of the coarse-grained model in Ref. [25] have been shown capable to describe the Statherin complexes at lower concentration and provide extra insight regarding the structure of the complexes, as well as aiding in explaining the effect of external conditions on the self-association, in terms of a balance between different interactions and entropy. The findings are summarized in Fig. 12. Hydrophobic interaction is shown to be the major driving force for the self-association, due to urea inhibiting complex formation. The size decrease as a result of increased temperature is regarded as an entropic effect, while electrostatic interactions were

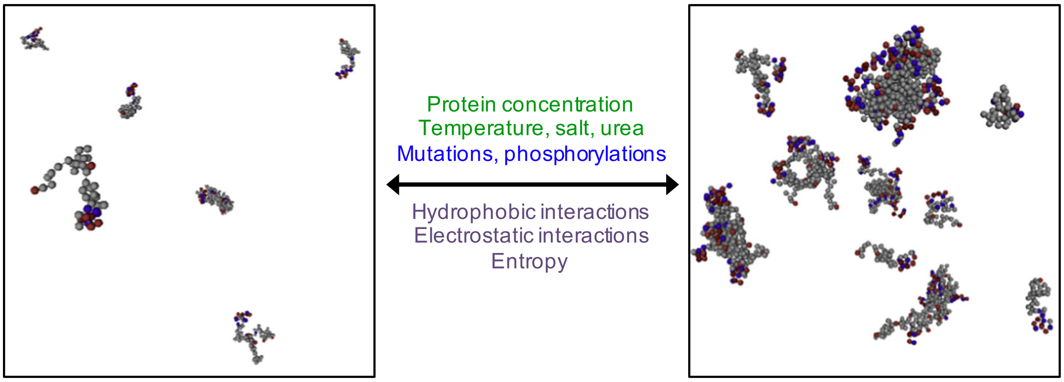

Fig. 12. Summary of what was shown to affect the Statherin association state. External factors are printed in green, chain characteristics in blue, and energetic and entropic factors in purple. In the snapshots, gray beads represent neutral residues; blue, positively charged residues; and red, negatively charged residues. The phosphorylated serines are marked in dark red. Counterions are omitted for clarity.

still shown to be of importance by balancing the hydrophobic attraction. In addition, it was demonstrated that mutations affecting the charge distribution can have a major effect on the self-association.

The self-association of Statherin is only one example of an IDP system dominated by intermolecular attractions; however, the similarities to micelle formation suggest that the established interactions are common for many systems, although with varying balance. It is therefore of interest to apply this model to other interacting IDPs in the future, as well as to continue the development for studies of systems with a higher complexity. Computational studies of IDP systems are advantageous in that it allows for separation of different contributions and a faster screening of mutations. In combination with experiments, it opens up for a deeper understanding of the function and behavior of IDPs.

Methods and Model

SAXS

Sample preparation

The buffers, all containing 20 mM Tris [>99.9%, CAS (77-86-1); Saveen Werner AB], and varying concentrations of NaCl [reagent grade, CAS (7647-14-5); Sharlau] and urea [ReagentPlus ≥99.5%, CAS (57-13-6); Sigma-Aldrich] were prepared with Milli-Q water, and by dropwise addition of 1 M HCl, the pH was set at room temperature to correspond to 8.1 at the measuring temperature. Thereafter, the buffers were filtered through a hydrophilic polypropylene 0.2 μm membrane (Pall Corporation).The Statherin powder (purchased from Genemed Synthesis, Inc.) was dissolved in buffer with a small addition of NaOH to increase the pH, since the protein powder contained trifluoroacetate. Concentrating cells (Vivaspin 2, 2000 MWCO, Prod. No. VS02H92; Sartorius, Cambridge, United Kingdom) were used to remove low-molecular-weight impurities. The samples were rinsed with buffer corresponding to 30 times the sample volume, by centrifugation at 358*g* at 8 °C. To ensure an exact background in the SAXS measurements, the samples were dialyzed (Slide-A-Lyzer Dialysis Cassette, 2000 MWCO, Prod. No. 66203 or Slide-A-Lyzer MINI Dialysis Unit, 2000 MWCO, Prod. No. 69580; Thermo Scientific, USA) overnight at 6 °C. Before the SAXS measurements, the samples were centrifuged at 18,400*g* at 6 °C for at least 2 h to remove impurities. Thereafter, they were diluted to a concentration series, and the protein concentration was determined with a nanodrop spectrometer using $\lambda = 280$ nm and $\varepsilon = 8740\ \mathrm{M}^{-1}\ \mathrm{cm}^{-1}$. The samples were centrifuged in small PCR tubes imminent to the SAXS measurements to remove any bubbles.

Measurements and analysis

SAXS experiments were performed at BM29, ESRF-Grenoble, France. The incident beam wavelength was 0.99 Å, and the distance between sample and detector (PILATUS 1M) was set to 2867 mm, giving the scattering vector 0.0039 – 0.49 Å^{-1}. The scattering vector, q, is defined as $q = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and λ is the wavelength of the incident beam. Several successive frames of the scattering from the samples were recorded with an exposure time of 0.5 or 1 s, depending on concentration and system. The scattering from the pure solvent, which was measured before and after each sample for the same exposure times, was subtracted from the sample scattering. Measurements were performed at 10, 20, 37 and 50 °C at 150 mM NaCl, and the forward scattering, I_0, was converted to absolute scale by water calibration. At 20 °C measurements were also performed for 10, 60 and 300 mM NaCl and 4 and 8 M urea. The data were processed and analyzed using the ATSAS package [45]. Special attention was paid to radiation damage by comparing the successive frames prior to background subtraction, and any affected data were rejected from further analysis. Both I_0 and R_g were determined from $P(r)$, although the Guinier approach was also used for comparison. The molecular weight used for calculating the association number was determined from I_0 (see Supplemental information). Considering standard uncertainties of the used values, the uncertainty of the association number can be estimated as approximately 10% [43,46].

For a description of the SEC inline with SAXS, used for obtaining the form factor of monomeric Statherin, we refer to Ref. [25].

CD

Protein was dissolved in and purified with 20 mM phosphate buffer (sodium phosphate dibasic dihydrate [Reag. Ph. Eur., CAS (10028-24-7); Sigma-Aldrich] and sodium phosphate monobasic monohydrate [ACS reagent, CAS (10049-21-5); Sigma-Aldrich]) at pH 8, using a concentrating cell, as described for the SAXS samples. The protein was diluted to approximately 0.13 mg/mL using 20 mM phosphate buffer with 10 or 150 mM NaF [≥99%, CAS (7681-49-4); Sigma-Aldrich] and for the 150 mM NaF with 0–8 M urea [ReagentPlus ≥99.5%, CAS (57–13-6); Sigma-Aldrich]. The samples were filtered using a 0.22-μm Millex–GV filter (Merk Millipore Ltd). CD spectra between 190 and 260 nm at temperatures 4 – 60 °C were recorded on a JASCO J-715 instrument with a PTC-348WI Peltier type cell holder for temperature control, averaging over three spectra for each sample, using a quartz cuvette with a 1-mm path length (HellmaAnalytics) and 20-nm/min scanning speed, 2-s response time, 1-nm band width, and 100-mdeg sensitivity. At 20 °C, further measurements were performed for samples with 150 mM NaF and 2–8 M urea. The ellipticity reported is the mean residue ellipticity, defined as

$$[\theta]_{\mathrm{MRW}} = \theta \cdot \mathrm{MRW}/(10 \cdot d \cdot c), \tag{1}$$

where θ is the observed ellipticity (mdeg), d the path length of the cell (cm), and c the protein concentration (mg/mL). The mean residue weight, MRW, is the molecular weight (Da) divided by the number of peptide bonds. The spectra were smoothed using a Savitzky–Golay filter. The effect of the Savitzky–Golay filter is presented in Fig. S4 in Supplemental information.

Coarse-grained model

We have employed a coarse-grained model in which each amino acid is modeled as a hard sphere, further described in Ref. [25]. For the inclusion of hydrophobic interaction, a short-ranged potential is added to the model:

$$U_{\mathrm{hphob}} = -\sum_{\mathrm{neutral}} \frac{\varepsilon_{\mathrm{hphob}}}{r_{ij}^{6}} \tag{2}$$

where the summation extends over all neutral amino acids, $r_{ij} = |\mathbf{R}_i - \mathbf{R}_j|$ is the center-to-center distance between two beads and $\mathbf{R}$ refers to the coordinate vector. $\varepsilon_{\mathrm{hphob}}$ is $1.32 \cdot 10^4$ kJ Å/mol, which corresponds to an attraction of 1.32 kT at closest contact, determined by comparing the average complex size with experimental results on the reference system.

Simulation aspects

The equilibrium properties of the model systems were obtained by Metropolis Monte Carlo simulations in the canonical (NVT) ensemble, utilizing the simulation package Molsim [47], version 4.8.8. Forty-five protein chains were enclosed in a cubic box of varying volume, dependent on the protein concentration. Periodic boundary conditions were applied in all directions. The long-ranged Coulomb interactions were truncated using the minimum image convention.

To accelerate the examination of the configurational space, five different types of displacements were allowed: (i) translational displacement of a single bead, (ii) pivot rotation [48,49], (iii) translation of the entire chain, (iv) slithering move [50], and (v) cluster displacements. Counterions were only moved individually by translation. The cluster displacement was performed as a translational displacement of the chain of a selected particle as well as all chains whose center of mass were less than 40 Å away from the selected particle. The cluster displacement was automatically rejected if the number of particles within the cluster changed, that is, if the displacement caused two clusters to merge. The probability of the different trial moves was weighted so that 80% of the trial moves were single bead displacements, 5% were pivot rotations, 5% were chain displacements, 3% were slithering moves, and 7% were cluster moves. Initially, the proteins were randomly placed in the box and an equilibrium simulation of typically $3 \cdot 10^5$ trial moves/bead was performed. The proceeding production run comprised at least 10^6 passes divided into subdivisions of 10^5 passes. To ensure accurately sampled simulations, the contact probability of each chain individually and the variations of contact number along the propagation of the simulation were analyzed (data not shown).

For all simulated quantities except the average association number, the reported uncertainty is one standard deviation of the mean. It is estimated from the deviation among the means of the subdivisions of the total number of MC passes, according to

$$\sigma^2(\langle x \rangle) = \frac{1}{n_s(n_s-1)} \sum_{s=1}^{n_s} \left(\langle x \rangle_s - \langle x \rangle \right)^2, \tag{3}$$

where $\langle x \rangle_s$ is the average of quantity x from one subdivision, $\langle x \rangle$ the average of x from the total simulation, and n_s the number of subdivisions. For the average association number, the reported uncertainty is the standard deviation of the means of all subdivisions.

Analyses

The calculation of the scattering profile from simulation is described in Ref. [25]. In the analyses of complexes, two chains were assigned to the same complex if the center-to-center distance between two beads in the two different chains was less than 5 Å. The same geometric condition was used for defining if a bead was in contact with another chain, which was the basis for monitoring the variations of contact number along the propagation, and calculating the contact probability for beads along the chain. Contact probability for the beads is defined as the number of passes in which the bead is in contact with at least one bead from another chain, divided by the total number of passes in the simulation. Similarly, contact probability for a chain is calculated as the number of passes in which the chain is in a complex divided by the total number of passes in the simulation and the overall contact probability is the average over all chains. The complex size probability distribution was calculated according to

$$P_n = \frac{n \langle N_n^{\mathrm{complex}} \rangle}{\sum_n n \langle N_n^{\mathrm{complex}} \rangle}, \tag{4}$$

where $\langle N_n^{\mathrm{complex}} \rangle$ is the average number of complexes consisting of n chains, and $\sum_n n \langle N_n^{\mathrm{complex}} \rangle$ is equal to the number of chains in the system, due to chain conservation. Note that P_n is weighted by the number of chains in a complex. The average association number was calculated from the complex size probability distribution, as

$$N_{\mathrm{assoc}} = \sum_n n P_n. \tag{5}$$

The radial number density profile was calculated for each complex size and bead type individually. The radial number density at each distance is defined as the number of beads within a shell at that distance from the center-of-mass of the complex core, divided by the shell volume. The complex core was defined to consist of the beads 15–44 in each chain.

The shape of the complexes was quantified by the principal moments of the gyration tensor and the asphericity. The gyration tensor was defined as

$$S = \frac{1}{N} \begin{pmatrix} \sum_i^N X_i^2 & \sum_i^N X_i Y_i & \sum_i^N X_i Z_i \\ \sum_i^N X_i Y_i & \sum_i^N Y_i^2 & \sum_i^N Y_i Z_i \\ \sum_i^N X_i Z_i & \sum_i^N Y_i Z_i & \sum_i^N Z_i^2 \end{pmatrix}, \tag{6}$$

where $A_i = (a_i - a_{\mathrm{com}})$ for $a = x, y, z$, and N is the number of beads in the complex. Transformation to a principal axis system such that

$$S = \mathrm{diag}\left(R_1^2, R_2^2, R_3^2\right) \tag{7}$$

diagonalizes S and $R_1^2 \geq R_2^2 \geq R_3^2$ are the eigenvalues of S, also called the principal moments of the gyration tensor. In the simulations, the ensemble averages of the eigenvalues were calculated for each complex size separately. The asphericity, defined as

$$\alpha_s = \frac{\left(\langle R_1^2 \rangle - \langle R_2^2 \rangle\right)\left(\langle R_2^2 \rangle - \langle R_3^2 \rangle\right)\left(\langle R_3^2 \rangle - \langle R_1^2 \rangle\right)}{2\left(\langle R_1^2 \rangle + \langle R_2^2 \rangle + \langle R_3^2 \rangle\right)^2}, \tag{8}$$

ranges between 0 for a perfect sphere and 1 for a rod.

Acknowledgments

The authors thank the European Synchrotron Radiation Facility (ESRF) for providing beamtime, Dr. Bart Van Laer at ESRF for providing assistance in using beamline BM29, and Magnus Bergvall's foundation for financial support. The study was supported by the Science Faculty project grant program for research with neutrons and synchrotron light (Lund University Strategic funds for MAX-IV and European Spallation Source). The simulations were performed on resources provided by the Swedish National Infrastructure for Computing at the Center for scientific and technical computing at Lund University (Lunarc).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2018.11.027.

Received 10 September 2018;
Received in revised form 12 November 2018;
Accepted 28 November 2018
Available online 7 December 2018

Keywords:
intrinsically disordered proteins;
SAXS;
self-association;
Monte Carlo simulations;
coarse-graining

Abbreviations used:
IDPs, intrinsically disordered proteins; IDRs, intrinsically disordered regions; SAXS, small-angle X-ray scattering; SEC, size-exclusion chromatography; PPII, poly-proline II.

References

[1] P. Tompa, Intrinsically unstructured proteins, Trends Biochem. Sci. 27 (10) (2002) 527–533, https://doi.org/10.1016/S0968-0004(02)02169-2.

[2] A. Dunker, J. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.M. Ratliff, K.W. Hipps, J. Ausio, M.S. Nissen, R. Reeves, C. Kang, C.R. Kissinger, R.W. Bailey, M.D. Griswold, W. Chiu, E.C. Garner, Z. Obradovic, Intrinsically disordered protein, J. Mol. Graph. Model. 19 (1) (2001) 26–59, https://doi.org/10.1016/S1093-3263(00)00138-8.

[3] A.K. Dunker, J. Gough, Sequences and topology: intrinsic disorder in the evolving universe of protein structure, Curr. Opin. Struct. Biol. 21 (3) (2011) 379–381, https://doi.org/10.1016/j.sbi.2011.04.002.

[4] J. Habchi, P. Tompa, S. Longhi, V.N. Uversky, Introducing protein intrinsic disorder, Chem. Rev. 114 (13) (2014) 6561–6588, https://doi.org/10.1021/cr400514h.

[5] A.K. Dunker, P. Romero, Z. Obradovic, E.C. Garner, C.J. Brown, Intrinsic protein disorder in complete genomes, Genome Inform. 11 (2000) 161–171, https://doi.org/10.11234/gi1990.11.161.

[6] P. Romero, Z. Obradovic, C. Kissinger, J. Villafranca, E. Garner, S. Guilliot, A. Dunker, Thousands of proteins likely

to have long disordered regions, Pac. Symp. Biocomput. 3 (1998) 437–448.

[7] J. Ward, J. Sodhi, L. McGuffin, B. Buxton, D. Jones, Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, J. Mol. Biol. 337 (3) (2004) 635–645, https://doi.org/10.1016/j.jmb.2004.02.002.

[8] B. Xue, A.K. Dunker, V.N. Uversky, Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life, J. Biomol. Struct. Dyn. 30 (2) (2012) 137–149, https://doi.org/10.1080/07391102.2012.675145.

[9] J. Liu, J.R. Faeder, C.J. Camacho, Toward a quantitative theory of intrinsically disordered proteins and their function, Proc. Natl. Acad. Sci. U. S. A. 106 (47) (2009) 19819–19823, https://doi.org/10.1073/pnas.0907710106.

[10] V.N. Uversky, Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders, Front. Aging Neurosci. 7 (2015) 18, https://doi.org/10.3389/fnagi.2015.00018.

[11] S.F. Banani, H.O. Lee, A.A. Hyman, M.K. Rosen, Biomolecular condensates: organizers of cellular biochemistry, Nat. Rev. Mol. Cell Biol. 18 (2017) 285–298.

[12] S. Weber, C. Brangwynne, Inverse size scaling of the nucleolus by a concentration-dependent phase transition, Curr. Biol. 25 (5) (2015) 641–646, https://doi.org/10.1016/j.cub.2015.01.012.

[13] C.P. Brangwynne, C.R. Eckmann, D.S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Jülicher, A.A. Hyman, Germline p granules are liquid droplets that localize by controlled dissolution/condensation, Science 324 (5935) (2009) 1729–1732, https://doi.org/10.1126/science.1172046.

[14] L.-P. Bergeron-Sandoval, N. Safaee, S. Michnick, Mechanisms and consequences of macromolecular phase separation, Cell 165 (5) (2016) 1067–1079, https://doi.org/10.1016/j.cell.2016.05.026.

[15] A.L. Darling, Y. Liu, C.J. Oldfield, V.N. Uversky, Intrinsically disordered proteome of human membrane-less organelles, Proteomics 18 (5–6) (2018) 1700193, https://doi.org/10.1002/pmic.201700193.

[16] V.N. Uversky, Protein intrinsic disorder-based liquid–liquid phase transitions in biological systems: complex coacervates and membrane-less organelles, Adv. Colloid Interf. Sci. 239 (2017) 97–114, https://doi.org/10.1016/j.cis.2016.05.012.

[17] V.N. Uversky, Intrinsically disordered proteins in overcrowded milieu: membrane-less organelles, phase separation, and intrinsic disorder, Curr. Opin. Struct. Biol. 44 (2017) 18–30, https://doi.org/10.1016/j.sbi.2016.10.015.

[18] S. Rauscher, R. Pomès, Molecular simulations of protein disorder, Biochem. Cell Biol. 88 (2) (2010) 269–290, https://doi.org/10.1139/O09-169.

[19] V.M. Burger, T. Gurry, C.M. Stultz, Intrinsically disordered proteins: where computation meets experiment, Polymers 6 (10) (2014) 2684–2719, https://doi.org/10.3390/polym6102684.

[20] R.B. Best, W. Zheng, J. Mittal, Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association, J. Chem. Theory Comput. 10 (11) (2014) 5113–5124, https://doi.org/10.1021/ct500569b.

[21] S. Piana, A.G. Donchev, P. Robustelli, D.E. Shaw, Water dispersion interactions strongly influence simulated structural properties of disordered protein states, J. Phys. Chem. B 119 (16) (2015) 5113–5123, https://doi.org/10.1021/jp508971m.

[22] J. Henriques, C. Cragnell, M. Skepö, Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment, J. Chem. Theory Comput. 11 (7) (2015) 3420–3431, https://doi.org/10.1021/ct501178z.

[23] S. Rauscher, V. Gapsys, M.J. Gajda, M. Zweckstetter, B.L. de Groot, H. Grubmüller, Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment, J. Chem. Theory Comput. 11 (11) (2015) 5513–5524, https://doi.org/10.1021/acs.jctc.5b00736.

[24] J. Henriques, M. Skepö, Molecular dynamics simulations of intrinsically disordered proteins: on the accuracy of the tip4p-d water model and the representativeness of protein disorder models, J. Chem. Theory Comput. 12 (7) (2016) 3407–3415, https://doi.org/10.1021/acs.jctc.6b00429.

[25] C. Cragnell, E. Rieloff, M. Skepö, Utilizing coarse-grained modeling and Monte Carlo simulations to evaluate the conformational ensemble of intrinsically disordered proteins and regions, J. Mol. Biol. 430 (16) (2018) 2478–2492, https://doi.org/10.1016/j.jmb.2018.03.006.

[26] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1) (1982) 105–132, https://doi.org/10.1016/0022-2836(82)90515-0.

[27] Y. Lin, S.L. Currie, M.K. Rosen, Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs, J. Biol. Chem. 292 (46) (2017) 19110–19120, https://doi.org/10.1074/jbc.M117.800466.

[28] C. Pak, M. Kosno, A. Holehouse, S. Padrick, A. Mittal, R. Ali, A. Yunus, D. Liu, R. Pappu, M. Rosen, Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein, Mol. Cell 63 (1) (2016) 72–85, https://doi.org/10.1016/j.molcel.2016.05.042.

[29] R. Williams, Z. Obradovic, V. Mathura, W. Braun, E. Garner, J. Young, S. Takayama, C. Brown, A. Dunker, The protein non-folding problem: amino acid determinants of intrinsic order and disorder, Pac. Symp. Biocomput. 2001 (6) (2001) 89–100.

[30] G.A. Naganagowda, T.L. Gururaja, M.J. Levine, Delineation of conformational preferences in human salivary statherin by 1H, 31P NMR and CD studies: sequential assignment and structure-function correlations, J. Biomol. Struct. Dyn. 16 (1) (1998) 91–107, https://doi.org/10.1080/07391102.1998.10508230.

[31] G.A. Elgavish, D.I. Hay, D.H. Schlesinger, 1H and 31P nuclear magnetic resonance studies of human salivary statherin, Int. J. Pept. Protein Res. 23 (3) (1984) 230–234, https://doi.org/10.1111/j.1399-3011.1984.tb02714.x.

[32] G. Goobes, R. Goobes, W.J. Shaw, J.M. Gibson, J.R. Long, V. Raghunathan, O. Schueler-Furman, J.M. Popham, D. Baker, C.T. Campbell, P.S. Stayton, G.P. Drobny, The structure, dynamics, and energetics of protein adsorption—lessons learned from adsorption of statherin to hydroxyapatite, Magn. Reson. Chem. 45 (S1) (2007) S32–S47, https://doi.org/10.1002/mrc.2123.

[33] N. Ramasubbu, L.M. Thomas, K.K. Bhandary, M.J. Levine, Structural characteristics of human salivary statherin: a model for boundary lubrication at the enamel surface, Crit. Rev. Oral Biol. Med. 4 (3) (1993) 363–370, https://doi.org/10.1177/10454411930040031501.

[34] P.A. Raj, M. Johnsson, M.J. Levine, G.H. Nancollas, Salivary statherin. Dependence on sequence, charge, hydrogen bonding potency, and helical conformation for adsorption to hydroxyapatite and inhibition of mineralization, J. Biol. Chem. 267 (9) (1992) 5968–5976.

[35] S.J. Whittington, B.W. Chellgren, V.M. Hermann, T.P. Creamer, Urea promotes polyproline II helix formation: implications for protein denatured states, Biochemistry 44 (16) (2005) 6269–6275, https://doi.org/10.1021/bi050124u.

[36] M. Kjaergaard, A.-B. Nørholm, R. Hendus-Altenburger, S.F. Pedersen, F.M. Poulsen, B.B. Kragelund, Temperature-dependent structural changes in intrinsically disordered proteins: formation of α-helices or loss of polyproline II? Protein Sci. 19 (8) (2010) 1555–1564, https://doi.org/10.1002/pro.435.

[37] M. Abu-Hamdiyyah, The effect of urea on the structure of water and hydrophobic bonding, J. Phys. Chem. 69 (8) (1965) 2720–2725, https://doi.org/10.1021/j100892a039.

[38] W. Bruning, A. Holtzer, The effect of urea on hydrophobic bonds: the critical micelle concentration of *n*-dodecyltrimethylammonium bromide in aqueous solutions of urea1, J. Am. Chem. Soc. 83 (23) (1961) 4865–4866, https://doi.org/10.1021/ja01484a044.

[39] U. Thapa, K. Ismail, Urea effect on aggregation and adsorption of sodium dioctylsulfosuccinate in water, J. Colloid Interface Sci. 406 (2013) 172–177, https://doi.org/10.1016/j.jcis.2013.06.009.

[40] J. Broecker, S. Keller, Impact of urea on detergent micelle properties, Langmuir 29 (27) (2013) 8502–8510, https://doi.org/10.1021/la4013747.

[41] A. Malliaris, J. Le Moigne, J. Sturm, R. Zana, Temperature dependence of the micelle aggregation number and rate of intramicellar excimer formation in aqueous surfactant solutions, J. Phys. Chem. 89 (12) (1985) 2709–2713, https://doi.org/10.1021/j100258a054.

[42] R. Zana, C. Weill, Effect of temperature on the aggregation behaviour of nonionic surfactants in aqueous solutions, J. Phys. Lett. 46 (20) (1985) 953–960.

[43] C. Moitzi, I. Portnaya, O. Glatter, O. Ramon, D. Danino, Effect of temperature on self-assembly of bovine β-casein above and below isoelectric pH. Structural analysis by cryogenic-transmission electron microscopy and small-angle x-ray scattering, Langmuir 24 (7) (2008) 3020–3029, https://doi.org/10.1021/la702802a.

[44] M. Kenward, M.D. Whitmore, A systematic Monte Carlo study of self-assembling amphiphiles in solution, J. Chem. Phys. 116 (8) (2002) 3455–3470, https://doi.org/10.1063/1.1445114.

[45] D. Franke, M.V. Petoukhov, P.V. Konarev, A. Panjkovich, A. Tuukkanen, H.D.T. Mertens, A.G. Kikhney, N.R. Hajizadeh, J.M. Franklin, C.M. Jeffries, D.I. Svergun, *ATSAS 2.8*: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions, J. Appl. Crystallogr. 50 (4) (2017) 1212–1225, https://doi.org/10.1107/S1600576717007786.

[46] D. Orthaber, A. Bergmann, O. Glatter, SAXS experiments on absolute scale with Kratky systems using water as a secondary standard, J. Appl. Crystallogr. 33 (2) (2000) 218–225, https://doi.org/10.1107/S0021889899015216.

[47] J. Reščič, P. Linse, MOLSIM: a modular molecular simulation software, J. Comput. Chem. 36 (16) (2015) 1259–1274, https://doi.org/10.1002/jcc.23919.

[48] M. Lal, Monte Carlo computer simulation of chain molecules. I, Mol. Phys. 17 (1) (1969) 57–64.

[49] N. Madras, A.D. Sokal, The pivot algorithm: a highly efficient Monte Carlo method for the self-avoiding walk, J. Stat. Phys. 50 (1) (1988) 109–186, https://doi.org/10.1007/BF01022990.

[50] F.T. Wall, F. Mandel, Macromolecular dimensions obtained by an efficient Monte-Carlo method without sample attrition, J. Chem. Phys. 63 (11) (1975) 4592–4595.