Improving the global dimensions of intrinsically disordered proteins in Martini 3

F. Emil Thomasen[1], **Francesco Pesce**[1], **Mette Ahrensback Roesgaard**[1], **Giulio Tesei**[1], **Kresten Lindorff-Larsen**[1*]

***For correspondence:** lindorff@bio.ku.dk (KLL)

[1] Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark

Abstract

Coarse-grained molecular dynamics simulations are a useful tool to determine conformational ensembles of intrinsically disordered proteins (IDPs). Here, we show that the coarse-grained force field Martini 3 underestimates the global dimensions of IDPs when compared with small angle X-ray scattering (SAXS) data. Increasing the strength of protein-water interactions favors more expanded conformations, improving agreement with SAXS data and alleviating problems with overestimated IDP-IDP interactions.

Introduction

Intrinsically disordered proteins (IDPs) are proteins that do not fold into a single well-defined structure, but rather sample a range of conformations (***Wright and Dyson, 1999***). Molecular dynamics (MD) simulations are a useful tool for structural characterization of IDPs. Using integrative methods, MD simulations can be used to determine conformational ensembles of IDPs in accordance with experimental data. Successful application of MD simulations relies on accurate force fields and adequate sampling of protein conformations (***Bottaro and Lindorff-Larsen, 2018***).

Coarse-grained MD simulations, where groups of atoms are represented by single beads, allow for efficient sampling of IDP conformations (***Ingólfsson et al., 2014***). One of the most widely used coarse-grained force fields for biomolecular systems is Martini (***Marrink et al., 2007***; ***Monticelli et al., 2008***). Martini maps two to four non-hydrogen atoms to one bead and is mainly parameterized against thermodynamic partitioning data. While Martini has been used successfully to study a wide range of biomolecular systems, earlier versions of the force field have been found to underestimate the global dimensions of flexible multidomain proteins (***Larsen et al., 2020***; ***Martin et al., 2021***) and overestimate protein-protein interactions (***Stark et al., 2013***; ***Berg and Peter, 2019***; ***Alessandri et al., 2019***; ***Benayad et al., 2021***; ***Majumder and Straub, 2021***; ***Lamprakis et al., 2021***). In order to favor more expanded conformations of multidomain proteins, we have previously used an approach based on increasing the strength of non-bonded interactions between protein and water beads (***Larsen et al., 2020***; ***Martin et al., 2021***), improving the agreement with SAXS experiments. Similarly, others have decreased the strength of non-bonded interactions between protein beads to improve the accuracy of IDP phase partitioning (***Benayad et al., 2021***) and protein-protein interactions (***Stark et al., 2013***).

A new version of the Martini force field, Martini 3, was recently released, featuring a rebalancing of non-bonded interaction terms and addition of new bead-types (***Souza et al., 2021***). Martini 3

shows improved accuracy for a wide range of systems in biology and materials science and a high level of transferability. Improved areas include molecular packing, transmembrane helix interactions, protein aggregation, and DNA base-pairing (***Souza et al., 2021***; ***Lamprakis et al., 2021***). Here, we have tested the ability of Martini 3 to reproduce the global dimensions of IDPs. We find that simulations with Martini 3 on average underestimate the radius of gyration (R_g) by $\approx 30\%$, and suggest a rescaling factor for increased protein-water interactions that improves agreement with small angle X-ray scattering (SAXS) data.

Results and Discussion

We chose a set of twelve IDPs and three multidomain proteins to cover a range of different systems with available SAXS data (***Riback et al., 2017***; ***Cordeiro et al., 2019***; ***Mylonas et al., 2008***; ***Riback et al., 2017***; ***Ahmed et al., 2021***; ***Martin et al., 2020***; ***Johnson et al., 2017***; ***Gomes et al., 2020***; ***Kjaergaard et al., 2010***; ***Jephthah et al., 2019***; ***Fagerberg et al., 2020***; ***Sonntag et al., 2017***; ***Martin et al., 2021***) and ran MD simulations of each protein using the Martini 3 force field. For all proteins, we found that the ensembles generated with Martini 3 were too compact when comparing the average R_g from the simulations with values calculated from SAXS profiles using Guinier analyses. A direct comparison with the experimental SAXS data also revealed deviations beyond the level expected by experimental errors (figure 1).

For atomistic force-fields, it has previously been shown that increasing the protein-water interactions will favour expanded conformations of IDPs, resulting in more accurate global dimensions (***Best et al., 2014***). Inspired by this approach, we increased the strength of protein-water interactions by rescaling the non-bonded Lennard-Jones potentials between all protein and water beads by a rescaling factor, λ. For all proteins, increased protein-water interactions (λ>1) resulted in an increased R_g and improved agreement with SAXS data as measured by the reduced χ^2 (χ^2_r). To determine an optimal value of λ, we scanned six λ-values from 1.04 to 1.14 for each protein. Based on the χ^2_r to SAXS data and agreement between R_g calculated from ensemble coordinates and R_g calculated from experimental SAXS profiles we chose λ=1.08 as the optimal value (figure 1a-c). We performed the same analysis for three multidomain proteins with flexible linkers, which all had an optimal λ around 1.04 (figure 1d-e), suggesting that the optimal value is different for folded domains and IDPs.

To further investigate the effect of rescaling protein-water interactions, we performed a number of tests comparing the original force field (λ=1) to the force field with increased protein-water interactions (λ=1.06 and 1.08). First, we tested the effect on the intrachain interactions in IDPs by comparing paramagnetic relaxation enhancement (PRE) data calculated from simulations of α-Synuclein, the FUS low-complexity domain (LCD) and the hnRNPA2 (A2) LCD to PRE experiments (***Dedmon et al., 2005***; ***Monahan et al., 2017***; ***Ryan et al., 2018***). We found that, for α-Synuclein, λ=1.06 and 1.08 improved the agreement with experimental PREs, while the agreement worsened for the two other proteins (figure 2a).

Next, we tested the effect of rescaling protein-water interactions on interchain IDP-IDP interactions. We simulated two copies of the FUS LCD at conditions matching interchain PRE experiments (***Monahan et al., 2017***) and calculated interchain PREs from the simulations for comparison. The original force-field did not show good agreement with experimental PREs, especially for the N-terminal spin-label at residue 16. However, the agreement worsened with λ=1.06 and 1.08 (figure 2b).

As a negative test of IDP-IDP interactions, we simulated two copies of α-Synclein, which should not interact under the given conditions based on PRE experiments (***Dedmon et al., 2005***). The original force field greatly overestimated the interaction of the two α-Synclein chains, predicting a 75±6 % population of the bound state. Increasing protein-water interactions by λ=1.06 and 1.08 reduced the population of the bound state to 12±1.3 % and 8±0.5 % respectively, thus improving the agreement with experiment (figure 2c). For comparison, we also calculated the population of the bound state in our simulations of the FUS LCD dimer, which should associate to a measurable

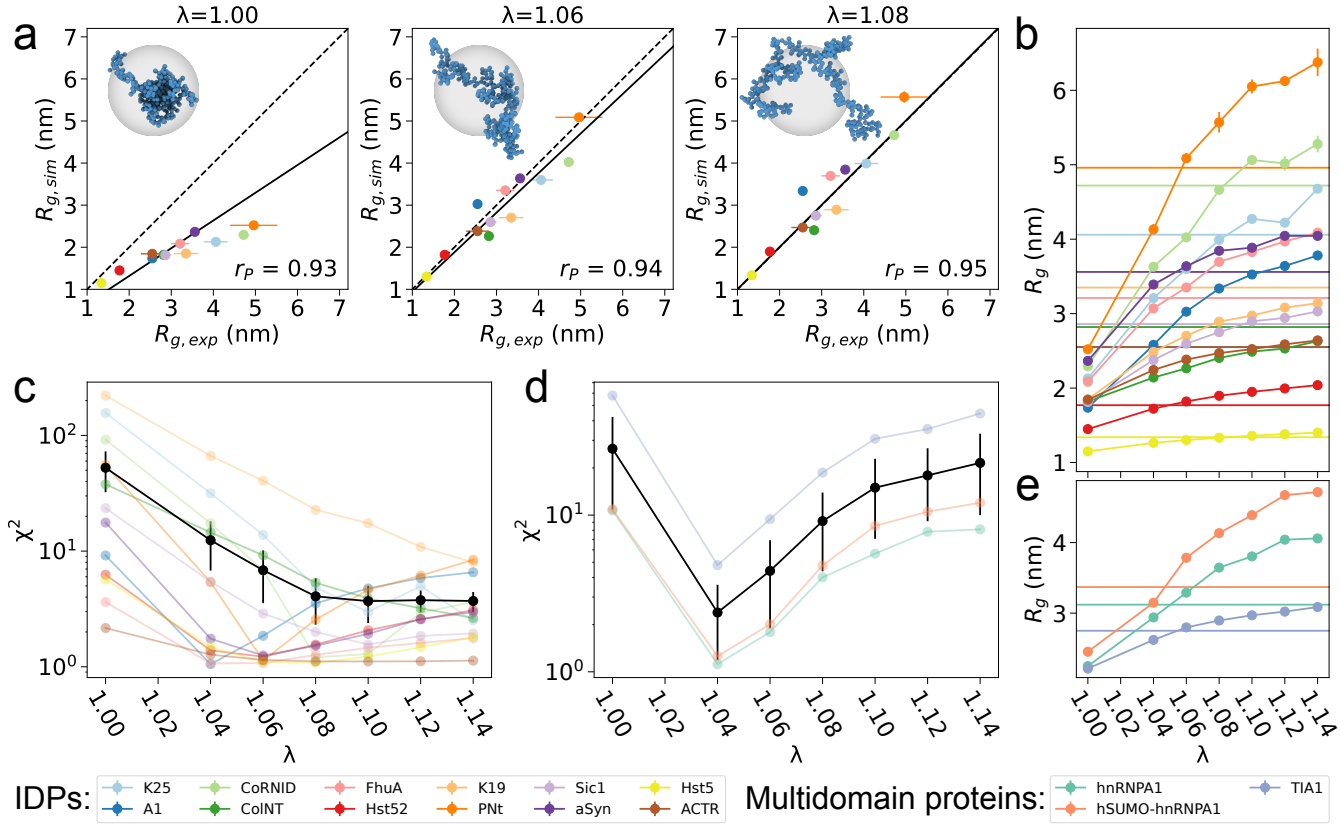

Figure 1. Increased protein-water interactions improve the agreement with SAXS data for IDPs and multidomain proteins
a. Average R_g from MD simulations with three different rescaling factors for the protein-water interactions (λ) plotted against experimental R_g from Guinier analysis of SAXS data for a set of twelve IDPs. Error bars for the experimental values were determined in the Guinier fit, and those for the simulations (here and elsewhere) were determined by block error analysis (***Flyvbjerg and Petersen, 1989***). Linear fit with intercept 0 weighted by experimental errors is shown as a solid line. Pearson correlation coefficient (r_P) is shown. The insert shows structures of Tau K25 (***Mylonas et al., 2008***) with the average R_g found for each λ. **b.** Average R_g from MD simulations over a range of λ-values for a set of twelve IDPs. Experimental values from Guinier analysis of SAXS data are shown as horizontal lines. **c–d.** Reduced χ_r^2 between SAXS profiles calculated from MD simulations and experimental SAXS profiles for a range of λ-values for a set of twelve IDPs (c) and three multidomain proteins (d). Average χ_r^2 is shown in black with standard error of the mean as error bars (note the log scale). **e.** Average R_g from MD simulations over a range of λ-values for three multidomain proteins. Experimental values from Guinier analysis of SAXS data are shown as horizontal lines. Data and scripts are available via github.com/KULL-Centre/papers/tree/main/2021/Martini-Thomasen-et-al

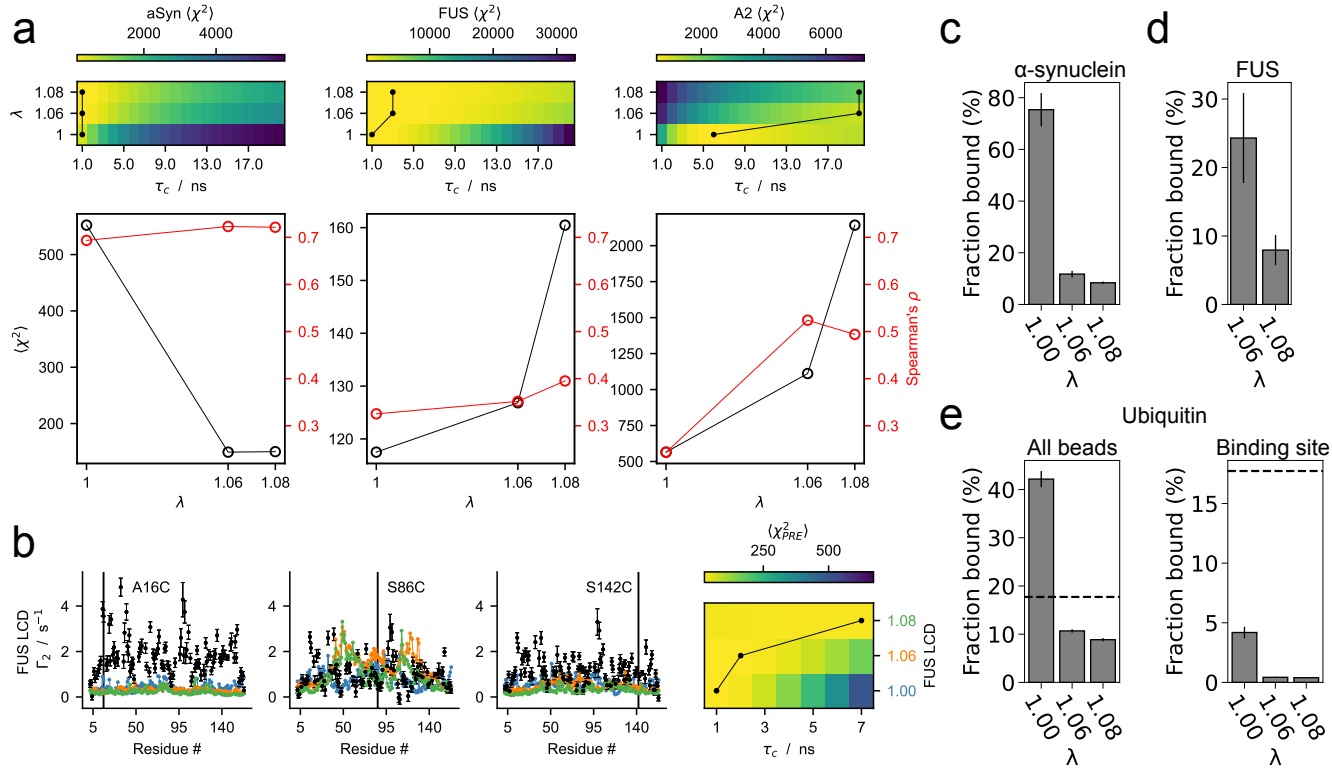

Figure 2. Effect of increased protein-water interactions on intrachain contacts and protein-protein interactions
a. Agreement between intrachain PREs calculated from MD simulations with different protein-water interaction rescaling factors λ and experimental PREs for the three IDPs α-synuclein, FUS LCD and hnRNPA2 LCD. Agreement is measured by χ_r^2 and Spearman correlation coefficient for PREs over all spin-label sites. Rotational correlation time τ_c was selected individually for each λ to minimize χ_r^2. **b.** Interchain PREs calculated from MD simulations with different λ-values of two copies of FUS LCD and comparison with experimental PREs (black). PREs are shown for three spin-label sites. Rotational correlation time τ_c was selected individually for each λ to minimize χ_r^2. **c–d.** Fraction bound calculated from MD simulations with different λ-values of two copies of α-synuclein (c) and FUS LCD (d). Error bars are standard error of the mean over ten replicas. **e.** Fraction bound calculated from MD simulations with different λ-values of two copies of ubiquitin. Definition of the bound state was based on the minimum distance between all beads (left) or the known binding site only (right) (***Liu et al., 2012***). Fraction bound based on experimentally determined K_d is shown as a dashed line (***Liu et al., 2012***). Error bars are standard error of the mean over ten replicas.

extent based on PRE experiments (***Monahan et al., 2017***). This analysis was only performed for simulations with λ=1.06 and 1.08, as we did not observe unbinding of the dimer after dimerization with the original force field (figure S3). In qualitative agreement with experiment, FUS had a higher population of the bound state than α-Synclein at λ=1.06, despite being at a four times lower concentration (figure 2c–d). The two proteins had similar population of the bound state at λ=1.08, indicating that a λ of 1.06 may be the optimal value for simulating IDP-IDP interactions.

Finally, we investigated the effect of rescaling protein-water interactions on interactions between folded proteins. Inspired by previous simulations (***Berg and Peter, 2019***) and NMR experiments (***Liu et al., 2012***), we simulated two copies of ubiquitin and calculated the population of the bound state. Simulations with Martini 3 appear to overestimate ubiquitin homodimerization when comparing the population of the bound state in the simulations with the value estimated by NMR chemical shift perturbations ($K_d = 4.9 \pm 0.3$ mM; ***Liu et al.*** (***2012***)). Because the interactions we observe in the simulations were not specific to the homodimerization site determined by NMR (***Liu et al., 2012***), this result was, however, dependent on whether the definition of the bound state was restricted to the known ubiquitin homodimerization site or not, illustrating that Martini 3 was not capturing the specificity of the interaction. However, λ=1.06 and 1.08 overly weakened the interaction, even when interactions outside the homodimerization site were considered part of the bound state (figure 2e).

Our results show that simulations with the Martini 3 force field result in too small global dimensions of IDPs and multidomain proteins, and that rescaling the Lennard-Jones potentials for protein-water interactions by a factor λ=1.08 improves agreement with experiments. For multidomain proteins containing flexible linkers or IDRs, a rescaling factor of λ=1.04 seems to be sufficient. Our results also show that Martini 3 greatly overestimates α-synuclein homodimerization, indicating that IDP-IDP interactions are too strong, but increasing protein-water interactions lead to a more accurate balance.

While increasing the strength of protein-water interactions improves the accuracy of the global dimensions of IDPs and the strength of IDP-IDP interactions, our results indicate that this is not the case for the interactions of folded proteins. First, increased protein-water interactions result in an underestimation of ubiquitin homodimerization. Second, multidomain proteins consisting of both folded domains and IDRs require a lower λ to reach agreement with SAXS data.

For the FUS LCD and the A2 LCD, the agreement with intrachain PREs are worsened by increasing λ to 1.06 and further worsened by going to 1.08. This indicates that, while the global dimensions of IDPs become more accurate with increased protein-water interactions, it potentially comes at the cost of specificity in intrachain interactions. In support of this, there is less sequence-dependent separation of R_g between different proteins with λ=1.08 than with the original force field (figure S2c). Additionally, the optimal value of λ correlates with the relative expansion of the IDP, showing that the optimal value of λ is partly sequence-dependent (figure 2Sb). Thus, a possible explanation for the worsened agreement with intrachain PREs for FUS LCD and A2 LCD is that they are relatively compact IDPs (***Ryan et al., 2018***), and λ values of 1.06 and 1.08 may be too high for these proteins. However, we also show evidence that the force-field is able to capture some sequence-specificity with λ=1.06: the FUS LCD self-associates more strongly than α-synuclein despite being at a four times lower concentration, in line with experimental observations (***Dedmon et al., 2005***; ***Monahan et al., 2017***). Taken together, these results suggest that a λ-value of 1.06 yields a good compromise between improving global dimensions and retaining specificity in interactions.

Alternatively, a λ-value can be chosen specifically for the system of interest if the level of compaction has been probed experimentally. This does, however, not necessarily entail optimizing λ-values for every condition of interest. For example, we have previously selected a single λ-value for simulations of hnRNPA1 (with a beta version of Martini 3) based on SAXS data at one salt concentration, and studied the effect of salt on the level of compaction by keeping the λ-value fixed but varying the salt concentration (***Martin et al., 2021***). A similar approach may be useful to transfer

λ-values between proteins with related sequence properties, for example in mutagenesis studies.

Conclusions

The functions of some IDPs and multidomain proteins depend on their ability to form biomolecular condensates (***Boeynaems et al., 2018***), often involving the formation of transient and multivalent protein-protein interactions and liquid-liquid phase separation (LLPS). Generally, the propensity of an IDP to undergo LLPS is correlated with its single-chain compactness (***Choi et al., 2020***). A modified version of Martini2.2 with decreased protein-protein interactions has already been shown to improve the description of LLPS of an IDP (***Benayad et al., 2021***), and Martini 3 has also been used to study salt-dependent condensate formation (***Tsanai et al., 2021***). We expect that increased protein-water interactions, yielding improved accuracy of the global dimensions of IDPs and weakened IDP-IDP interactions, will be useful in future applications of Martini 3 to study the role of IDPs in biomolecular condensates as well as their single-chain conformations and dynamics.

Acknowledgments

We thank Simone Orioli, Thea K. Schulze and Yong Wang for useful discussions and suggestions. We acknowledge the use of computational resources from Computerome 2.0. This research was supported by the Lundbeck Foundation BRAINSTRUC initiative (R155-2015-2666 to KLL) and the EU Horizon 2020 Marie Skłodowska-Curie grant agreement (101025063 to GT)

Methods

IDP simulations

We selected a set of twelve IDPs of varied sequence, with lengths between 24 and 334 amino acid residues and with SAXS data available: the N-terminal region of pertactin (PNt) (***Riback et al., 2017***), the NR interaction domain of N-CoR (CoRNID) (***Cordeiro et al., 2019***), two deletion mutants of Tau (K19 and K25) (***Mylonas et al., 2008***), the âĂIJplugâĂİ domain from a TonB-dependent receptor (FhuA) (***Riback et al., 2017***), α-synuclein (aSyn) (***Ahmed et al., 2021***), the low-complexity domain of hnRNPA1 (A1) (***Martin et al., 2020***), the T-domain of colicin N (ColNT) (***Johnson et al., 2017***), Sic1 (***Gomes et al., 2020***), the activation domain of ACTR (ACTR) (***Kjaergaard et al., 2010***), Histatin-5 (Hst5) (***Jephthah et al., 2019***) and a tandem repeat of Histatin-5 (Hst52) (***Fagerberg et al., 2020***).

We performed all MD simulations using Gromacs 2020.3 (***Abraham et al., 2015***) and the Martini3.0 force field (or adapted force fields with rescaled protein-water interactions) (***Souza et al., 2021***). Proteins were coarse-grained using the Martinize2 python script, placed in a dodecahedral box using Gromacs and solvated using the Insane python script (***Wassenaar et al., 2015***). Initial box size was chosen by using starting structures from simulations in ***Tesei et al.*** (***2021b***) corresponding to the 95th percentile of R_g-distributions and using Gromacs *editconf* with the flag *-d 4.0*. Box size was later increased if necessary. NaCl concentration was set to match the conditions in SAXS experiments and to neutralize the system. No secondary structure or elastic network model was assigned with Martinize2 for IDPs and IDRs. Energy minimization was performed using steepest descent for 10,000 steps with a 30 fs timestep. The Lennard-Jones potentials between all protein and water beads were rescaled by a factor λ. Seven values of λ were tested for each system: 1.00 (original force-field), 1.04, 1.06, 1.08, 1.10, 1.12 and 1.14. The systems were equilibrated for 10 ns with a 2 fs timestep using the Velocity-Rescaling thermostat (***Bussi et al., 2007***) and Parinello-Rahman barostat (***Parrinello and Rahman, 1981***). Production simulations were run for between 16 µs and 100 µs with a 20 fs timestep using the Velocity-Rescaling thermostat and Parinello-Rahman barostat. The temperature was set to match conditions in SAXS experiments and the pressure was set to 1 bar. Non-bonded interactions were treated with the Verlet cut-off scheme. The cut-off for Van der Waals interactions was set to 1.1 nm. Coulomb interactions were treated using the reaction-field method with a 1.1 nm cut-off and dielectric constant of 15. Frames were saved every 1 ns. Periodic boundary conditions were treated with Gromacs *trjconv* with the flags *-pbc whole*

-center. Simulation convergence was assessed using block-error analysis (***Flyvbjerg and Petersen, 1989***) of the R_g using the BLOCKING code (https://github.com/fpesceKU/BLOCKING). Simulations were backmapped to all-atom using a modified (***Larsen et al., 2020***) version of the Backward algorithm (***Wassenaar et al., 2014***), in which simulation runs are excluded and energy minimization is shortened to 200 steps.

We also ran MD simulations of two IDPs with paramagnetic relaxation enhancement (PRE) data available: the low-complexity domain of FUS (FUS) (***Monahan et al., 2017***) and low-complexity domain of hnRNPA2 (A2) (***Ryan et al., 2018***). For these proteins we set the NaCl concentration and temperature to match the conditions in PRE experiments. Additionally, we reran simulations of α-synclein at 283 K to match conditions in PRE experiments (***Dedmon et al., 2005***). The protocol was otherwise identical to above.

Multidomain protein simulations

We selected a set of three multidomain proteins with SAXS data available: full-length hnRNPA1 (hnRNPA1), full-length hnRNPA1 with an N-terminal His-SUMO tag (hSUMO-hnRNPA1) and TIA-1. SAXS data and initial structures of hnRNPA1 and hSUMO-hnRNPA1 were taken from ***Martin et al. (2021)***. These structures were built based on the structures of SUMO1 (PDB: 1A5R) (***Bayer et al., 1998***) and the RRM1 and RRM2 domains (PDB: 1HA1) (***Shamoo et al., 1997***). The initial structure of TIA-1 was taken from ***Larsen et al. (2020)*** and SAXS data was taken from ***Sonntag et al. (2017)***. The structure was built based on the structures of RRM1 (PDB 5O2V) (***Sonntag et al., 2017***), RRM2 (PDB: 5O3J) (***Sonntag et al., 2017***) and the RRM2-RRM3 complex (PDB: 2MJN) (***Wang et al., 2014***).

Simulations of multidomain proteins were set up and run using the same protocol as for the IDP simulations with a few exceptions: (i) Secondary structure was assigned with DSSP (***Kabsch and Sander, 1983***) in Martinize2. (ii) An elastic network model was applied with Martinize2 to keep folded domains intact. Interdomain elastic restraints and the elastic restraints in disordered regions and linker regions were removed. The elastic restraints consisted of a harmonic potential of 500 kJ mol^{-1} nm^{-2} between backbone beads within a 1.2 nm cut-off.

Ubiquitin dimerization simulations

Initial structures of ubiquitin were taken from ***Vijay-Kumar et al. (1987)*** (PDB: 1UBQ). Simulations of ubiquitin were set up and run using the same protocol as for IDP simulations with a few exceptions: (i) Two copies of ubiquitin were placed in a 14.92 nm x 14.92 nm x 14.92 nm cubic box, giving a protein concentration of 1 mM. (ii) Secondary structure was assigned with DSSP (***Kabsch and Sander, 1983***) in Martinize2. (iii) An elastic network model was applied with Martinize2. We removed elastic restraints from the C-terminus (residue 72-76) of ubiquitin to allow for flexibility (***Lindorff-Larsen et al., 2005***). The elastic restraints consisted of a harmonic potential of 500 kJ mol^{-1} nm^{-2} between backbone beads within a 0.9 nm cut-off. We ran simulations testing three different values of λ: 1.00, 1.06 and 1.08. For each value of λ, we ran ten replicas of 40 μs each.

FUS LCD and α-synuclein dimerization simulations

Simulations of two copies of FUS and two copies of α-synuclein were set up and run using the same protocol as for IDP simulations with a few exceptions: Two copies of FUS were placed in a 40.5 nm x 40.5 nm x 40.5 nm cubic box, giving a protein concentration of 50 μM to match PRE experiments. Two copies of α-synuclein were placed in a 25.51 nm x 25.51 nm x 25.51 nm cubic box, giving a protein concentration of 200 μM to match PRE experiments. We ran simulations testing three different values of λ: 1.00, 1.06 and 1.08, with ten replicas for each λ. For simulations of FUS at λ=1.00, replicas were run for between 11 and 13.5 μs each. However, we did not observe unbinding of the dimer after dimerization in any of the replicas, so these simulations were not extended further. For simulations of FUS at λ=1.06 and 1.08, replicas were run for between 25 and 29 μs each. For simulations of α-synclein, replicas were run for between 12 and 27 μs each.

Calculating the radius of gyration

We calculated the radius of gyration (R_g) from the coarse-grained trajectories using Gromacs *gyrate* (***Abraham et al., 2015***). Error bars on simulation R_g were determined using block-error analysis (***Flyvbjerg and Petersen, 1989***). Experimental R_g and corresponding error bars were calculated from SAXS profiles by Guinier analysis using ATSAS AUTORG with default settings (***Petoukhov et al., 2007***).

SAXS calculations

After each trajectory had been backmapped to all-atom resolution, we extracted 15000 frames (evenly distributed in the time-series) to calculate SAXS profiles using Pepsi-SAXS (***Grudinin et al., 2017***). To avoid potential problems of overfitting the parameters for the contrast of the hydration layer ($\delta\rho$) and the displaced solvent (r_0) (if these are fitted individually to each structure) we used values that have previously been shown to provide good agreement with experiment for flexible proteins (***Pesce and Lindorff-Larsen, 2021***). Values for the intensity of the forward scattering ($I(0)$) and the constant background (*cst*) were fitted globally with least-squares regression weighted by the experimental errors using the Scikit-learn python library (***Pedregosa et al., 2011***).

To quantify the agreement between experimental SAXS profiles and those calculated from simulations, we calculated the reduced χ^2:

$$\chi_r^2 = \frac{1}{m}\sum_q^m \frac{(I_q^{CALC} - I_q^{EXP})^2}{\sigma(BIFT)_q^2} \quad (1)$$

Here *m* is the number of data points, I_q^{CALC} and I_q^{EXP} are the averaged calculated SAXS intensity and the experimental SAXS intensity at scattering angle *q*, and $\sigma(BIFT)_q$ is the error for the experimental intensity at scattering angle *q* corrected according to: $\sigma(BIFT)_q = \sigma_q\sqrt{\chi^2_{r,BIFT}}$, where σ_q is the experimental error and $\chi^2_{r,BIFT}$ quantifies the agreement between the experimental SAXS data and the model SAXS profile calculated from the pair distance distribution function obtained through the Bayesian Indirect Fourier Transform algorithm (BIFT) (***Hansen, 2000***). This approach has been shown to lead to improved error estimates for experimental SAXS profiles (***Larsen and Pedersen, 2021***) and, here, made it possible to compare more directly and average over the χ_r^2 from the different systems. BIFT optimizes the hyperparameter D_{max} (maximum distance between scattering particles in the system); as an initial estimate of D_{max}, we used the D_{max} over all simulations for each protein.

PRE calculations

We used the DEER-PREdict software (***Tesei et al., 2021a***) to calculate PRE NMR data from all-atom backmapped trajectories. DEER-PREdict implements a model-free formalism (***Iwahara et al., 2004***) combined with a rotamer library approach to describe the MTSL spin-label probe (***Polyhach et al., 2011***). We assumed an effective correlation time of the spin label, τ_t, of 100 ps and fitted an overall molecular correlation time, τ_c, within the interval $1 \leq \tau_c \leq 20$ ns. Additionally, to calculate PRE intensity ratios, we assumed a transverse relaxation rate for the diamagnetic protein of 10 s^{-1} and approximated the total INEPT time of the HSQC measurement to 10 ms (***Battiste and Wagner, 2000***). We calculated intermolecular PRE rates from two-chain simulations treating one chain as spin-labeled and the other as ^{15}N-labeled. We averaged the PRE rates obtained for the two combinations of spin-labeled and ^{15}N-labeled chain, fitting τ_c to this average. Agreement between calculated and experimental PREs was quantified by the reduced χ^2 over all spin-label sites:

$$\chi_r^2 = \frac{1}{N_{labels}N_{res}}\sum_j^{N_{labels}}\sum_i^{N_{res}}\left(\frac{Y_{ij}^{exp} - Y_{ij}^{calc}}{\sigma_{ij}^{exp}}\right)^2 \quad (2)$$

Where N_{labels} and N_{res} are the number of spin-labels and residues, Y_{ij}^{exp} and Y_{ij}^{calc} are the experimental and calculated PRE rates for label j and residue i, and σ_{ij}^{exp} is the experimental error of the

PRE rate for label j and residue i.

Dimerization calculations

We analyzed the population of the bound and unbound states of ubiquitin, FUS and α-synuclein homodimers in our simulations. We calculated the minimum distance between any beads in the two proteins over the trajectory using Gromacs *mindist* (***Abraham et al., 2015***). The fraction bound was defined as the fraction of frames where the minimum distance was below 0.8 nm, and error bars as the standard error of the mean over the ten replica simulations. For simulations of ubiquitin, the fraction bound was also calculated using the minimum distance only between beads in the binding site (residue 8, 13, 44, 45, 46, 49, 67, 68, 70, 71, and 73) defined by NMR chemical shift perturbations (***Liu et al., 2012***). This greatly reduced population of the bound state, showing that Martini3 did not capture the specificity of the interaction.

Data availability

Scripts and data are at github.com/KULL-Centre/papers/tree/main/2021/Martini-Thomasen-et-al

Protein systems

Table 1. IDP simulations for SAXS and R_g calculations: Number of amino acid residues (N_R), box size (d), experimental R_g, simulation temperature (T), and salt concentration in the simulation (c_s).

Protein	N_R	d (nm)	SAXS R_g (nm)	T (K)	c_s (M)	SAXS ref.
Hst5	24	13.7	1.38 ± 0.01	293	0.15	***Jephthah et al.* (*2019*)**
$(Hst5)_2$	48	17.4	1.87 ± 0.07	298	0.15	***Fagerberg et al.* (*2020*)**
ACTR	71	18.9	2.63 ± 0.1	278	0.2	***Kjaergaard et al.* (*2010*)**
Sic1	92	21.4	3.00 ± 0.4	293	0.2	***Gomes et al.* (*2020*)**
ColNT	98	20.5	2.83 ± 0.1	277	0.4	***Johnson et al.* (*2017*)**
K19	99	20.4	3.50 ± 0.1	288	0.15	***Mylonas et al.* (*2008*)**
A1	137	21.5	2.72 ± 0.02	296	0.05	***Martin et al.* (*2020*)**
αSyn	140	24.1	3.55 ± 0.05	293	0.2	***Ahmed et al.* (*2021*)**
FhuA	144	23.9	3.34 ± 0.1	298	0.15	***Riback et al.* (*2017*)**
K25	185	27.4	4.10 ± 0.2	288	0.15	***Mylonas et al.* (*2008*)**
CoRNID	271	32.5	4.70 ± 0.2	293	0.2	***Cordeiro et al.* (*2019*)**
PNt	334	31.2	5.11 ± 0.2	298	0.15	***Riback et al.* (*2017*)**

Table 2. Multidomain protein simulations for SAXS and R_g calculations: Number of amino acid residues (N_R), box size (d), experimental R_g, simulation temperature (T), and salt concentration in the simulation (c_s).

Protein	N_R	d (nm)	SAXS R_g (nm)	T (K)	c_s (M)	SAXS ref.
TIA1	275	17.9	2.75 ± 0.031	300	0.1	***Sonntag et al.* (*2017*)**
A1	314	28.6	3.12 ± 0.022	300	0.15	***Martin et al.* (*2021*)**
hSUMO-A1	433	29.1	3.37 ± 0.014	300	0.1	***Martin et al.* (*2021*)**

Table 3. IDP simulations for single-chain PRE calculations: Number of amino acid residues (N_R), box size (d), experimental R_g, simulation temperature (T), and salt concentration in the simulation (c_s).

Protein	N_R	d (nm)	T (K)	c_s (M)	PRE ref.
αSyn	140	24.1	283	0.2	***Dedmon et al.* (*2005*)**
A2	155	21.8	298	0.005	***Ryan et al.* (*2018*)**
FUS	163	19.4	298	0.15	***Monahan et al.* (*2017*)**

Table 4. Protein dimerization simulations: Number of amino acid residues (N_R), box size (d), experimental R_g, simulation temperature (T), and salt concentration in the simulation (c_s).

Protein	N_R	d (nm)	T (K)	c_s (M)	PRE or affinity ref.
αSyn	140x2	25.5	283	0.125	***Dedmon et al.* (*2005*)**
FUS	163x2	40.5	298	0.15	***Monahan et al.* (*2017*)**
Ubq	76x2	14.9	303	0.11	***Liu et al.* (*2012*)**

References

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015; 1-2:19–25. doi: 10.1016/j.softx.2015.06.001.

Ahmed MC, Skaanning LK, Jussupow A, Newcombe EA, Kragelund BB, Camilloni C, Langkilde AE, Lindorff-Larsen K. Refinement of α-Synuclein Ensembles Against SAXS Data: Comparison of Force Fields and Methods. Frontiers in Molecular Biosciences. 2021; 8(April):1–13. doi: 10.3389/fmolb.2021.654333.

Alessandri R, Souza PC, Thallmair S, Melo MN, De Vries AH, Marrink SJ. Pitfalls of the Martini model. Journal of chemical theory and computation. 2019; 15(10):5448–5460.

Battiste JL, Wagner G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. Biochemistry. 2000; 39(18):5355–5365.

Bayer P, Arndt A, Metzger S, Mahajan R, Melchior F, Jaenicke R, Becker J. Structure determination of the small ubiquitin-related modifier SUMO-1. Journal of Molecular Biology. 1998; 280(2):275–286. doi: 10.1006/jmbi.1998.1839.

Benayad Z, Von Bülow S, Stelzl LS, Hummer G. Simulation of FUS Protein Condensates with an Adapted Coarse-Grained Model. Journal of Chemical Theory and Computation. 2021; 17(1):525–537. doi: 10.1021/acs.jctc.0c01064.

Berg A, Peter C. Simulating and analysing configurational landscapes of protein–protein contact formation. Interface focus. 2019; 9(3):20180062.

Best RB, Zheng W, Mittal J. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. Journal of Chemical Theory and Computation. 2014; 10(11):5113–5124. doi: 10.1021/ct500569b.

Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, Schymkowitz J, Shorter J, Wolozin B, Van Den Bosch L, Tompa P, Fuxreiter M. Protein Phase Separation: A New Phase in Cell Biology. Trends in Cell Biology. 2018; 28(6):420–435. http://dx.doi.org/10.1016/j.tcb.2018.02.004, doi: 10.1016/j.tcb.2018.02.004.

Bottaro S, Lindorff-Larsen K. Biophysical experiments and biomolecular simulations: A perfect match? Science. 2018 7; 361(6400):355 LP – 360. http://science.sciencemag.org/content/361/6400/355.abstract, doi: 10.1126/science.aat4010.

Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. Journal of Chemical Physics. 2007; 126(1):1–7. doi: 10.1063/1.2408420.

Choi JM, Holehouse AS, Pappu RV. Physical principles underlying the complex biology of intracellular phase transitions. Annual Review of Biophysics. 2020; 49:107–133.

Cordeiro TN, Sibille N, Germain P, Barthe P, Boulahtouf A, Allemand F, Bailly R, Vivat V, Ebel C, Barducci A, Bourguet W, le Maire A, Bernadó P. Interplay of Protein Disorder in Retinoic Acid Receptor Heterodimer and Its Corepressor Regulates Gene Expression. Structure. 2019; 27(8):1270–1285. doi: 10.1016/j.str.2019.05.001.

Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM. Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations. Journal of the American Chemical Society. 2005; 127(2):476–477. doi: 10.1021/ja044834j.

Fagerberg E, Månsson LK, Lenton S, Skepö M. The Effects of Chain Length on the Structural Properties of Intrinsically Disordered Proteins in Concentrated Solutions. Journal of Physical Chemistry B. 2020; 124(52):11843–11853. doi: 10.1021/acs.jpcb.0c09635.

Flyvbjerg H, Petersen HG. Error estimates on averages of correlated data. The Journal of Chemical Physics. 1989; 91(1):461–466. doi: 10.1063/1.457480.

Gomes GNW, Krzeminski M, Namini A, Martin EW, Mittag T, Head-Gordon T, Forman-Kay JD, Gradinaru CC. Conformational Ensembles of an Intrinsically Disordered Protein Consistent with NMR, SAXS, and Single-Molecule FRET. Journal of the American Chemical Society. 2020; 142(37):15697–15710. doi: 10.1021/jacs.0c02088.

Grudinin S, Garkavenko M, Kazennov A. Pepsi-SAXS: An adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles. Acta Crystallographica Section D: Structural Biology. 2017; 73(5):449–464. doi: 10.1107/S2059798317005745.

Hansen S. Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering. Journal of Applied Crystallography. 2000 12; 33(6):1415–1421. https://doi.org/10.1107/S0021889800012930, doi: 10.1107/S0021889800012930.

Ingólfsson HI, Lopez CA, Uusitalo JJ, de Jong DH, Gopal SM, Periole X, Marrink SJ. The power of coarse graining in biomolecular simulations. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2014; 4(3):225–248. doi: 10.1002/wcms.1169.

Iwahara J, Schwieters CD, Clore GM. Ensemble Approach for NMR Structure Refinement against 1H Paramagnetic Relaxation Enhancement Data Arising from a Flexible Paramagnetic Group Attached to a Macromolecule. J Am Chem Soc. 2004 4; 126(18):5879–5896.

Jephthah S, Staby L, Kragelund BB, Skepö M. Temperature Dependence of Intrinsically Disordered Proteins in Simulations: What are We Missing? Journal of Chemical Theory and Computation. 2019; 15(4):2672–2683. doi: 10.1021/acs.jctc.8b01281.

Johnson CL, Solovyova AS, Hecht O, Macdonald C, Waller H, Grossmann JG, Moore GR, Lakey JH. The Two-State Prehensile Tail of the Antibacterial Toxin Colicin N. Biophysical Journal. 2017; 113(8):1673–1684. doi: 10.1016/j.bpj.2017.08.030.

Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983 12; 22(12):2577–2637. https://doi.org/10.1002/bip.360221211, doi: https://doi.org/10.1002/bip.360221211.

Kjaergaard M, Nørholm AB, Hendus-Altenburger R, Pedersen SF, Poulsen FM, Kragelund BB. Temperature-dependent structural changes in intrinsically disordered proteins: Formation of α-helices or loss of polyproline II? Protein Science. 2010; 19(8):1555–1564. doi: 10.1002/pro.435.

Lamprakis C, Andreadelis I, Manchester J, Velez-Vega C, Duca JS, Cournia Z. Evaluating the efficiency of the Martini force field to study protein dimerization in aqueous and membrane environments. Journal of Chemical Theory and Computation. 2021; 17(5):3088–3102.

Larsen AH, Pedersen MC. Experimental noise in small-angle scattering can be assessed using the Bayesian indirect Fourier transformation. Journal of Applied Crystallography. 2021 10; 54(5). https://doi.org/10.1107/S1600576721006877, doi: 10.1107/S1600576721006877.

Larsen AH, Wang Y, Bottaro S, Grudinin S, Arleth L, Lindorff-Larsen K. RESEARCH ARTICLE Combining molecular dynamics simulations with small-angle X-ray and neutron scattering data to study multi-domain proteins in solution. PLoS Computational Biology. 2020; 16(4):1–29. http://dx.doi.org/10.1371/journal.pcbi.1007870, doi: 10.1371/journal.pcbi.1007870.

Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M. Simultaneous determination of protein structure and dynamics. Nature. 2005; 433(7022):128–132. https://doi.org/10.1038/nature03199, doi: 10.1038/nature03199.

Liu Z, Zhang WP, Xing Q, Ren X, Liu M, Tang C. Noncovalent dimerization of ubiquitin. Angewandte Chemie - International Edition. 2012; 51(2):469–472. doi: 10.1002/anie.201106190.

Majumder A, Straub JE. Addressing the Excessive Aggregation of Membrane Proteins in the MARTINI Model. Journal of Chemical Theory and Computation. 2021; 17(4):2513–2521.

Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH. The MARTINI force field: Coarse grained model for biomolecular simulations. Journal of Physical Chemistry B. 2007; 111(27):7812–7824. doi: 10.1021/jp071097f.

Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, Grace CR, Soranno A, Pappu RV, Mittag T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science. 2020; 367(6478):694–699. doi: 10.1126/science.aaw8653.

Martin EW, Thomasen FE, Milkovic NM, Cuneo MJ, Grace CR, Nourse A, Lindorff-Larsen K, Mittag T. Interplay of folded domains and the disordered low-complexity domain in mediating hnRNPA1 phase separation. Nucleic Acids Research. 2021; 49(5):2931–2945. doi: 10.1093/nar/gkab063.

Monahan Z, Ryan VH, Janke AM, Burke KA, Rhoads SN, Zerze GH, O'Meally R, Dignon GL, Conicella AE, Zheng W, Best RB, Cole RN, Mittal J, Shewmaker F, Fawzi NL. Phosphorylation of the FUS lowâĂŘcomplexity domain disrupts phase separation, aggregation, and toxicity . The EMBO Journal. 2017; 36(20):2951–2967. doi: 10.15252/embj.201696394.

Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI coarse-grained force field: Extension to proteins. Journal of Chemical Theory and Computation. 2008; 4(5):819–834. doi: 10.1021/ct700324x.

Mylonas E, Hascher A, Bernadó P, Blackledge M, Mandelkow E, Svergun DI. Domain conformation of tau protein studied by solution small-angle X-ray scattering. Biochemistry. 2008; 47(39):10345–10353. doi: 10.1021/bi800900d.

Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics. 1981; 52(12):7182–7190. doi: 10.1063/1.328693.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay Ã. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011; 12:2825–2830.

Pesce F, Lindorff-Larsen K. Refining conformational ensembles of flexible proteins against small-angle X-ray scattering data. bioRxiv. 2021 1; p. 2021.05.29.446281. http://biorxiv.org/content/early/2021/09/09/2021.05.29.446281.abstract, doi: 10.1101/2021.05.29.446281.

Petoukhov MV, Konarev PV, Kikhney AG, Svergun DI. ATSAS 2.1 towards automated and web-supported small-angle scattering data analysis. Journal of Applied Crystallography. 2007 4; 40(s1):s223–s228. https://doi.org/10.1107/S0021889807002853, doi: 10.1107/S0021889807002853.

Polyhach Y, Bordignon E, Jeschke G. Rotamer libraries of spin labelled cysteines for protein studies. Phys Chem Chem Phys. 2011; 13(6):2356–2366. https://doi.org/10.1039/c0cp01865a.

Riback JA, Bowman MA, Zmyslowski AM, Knoverek CR, Jumper JM, Hinshaw JR, Kaye EB, Freed KF, Clark PL, Sosnick TR. Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. Science (New York, NY). 2017; 358(6360):238–241. http://www.ncbi.nlm.nih.gov/pubmed/29026044.

Ryan VH, Dignon GL, Zerze GH, Chabata CV, Silva R, Conicella AE, Amaya J, Burke KA, Mittal J, Fawzi NL. Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. Molecular cell. 2018 2; 69(3):465–479. doi: 10.1016/j.molcel.2017.12.022.

Shamoo Y, Krueger U, Rice LM, Williams KR, Steitz TA. Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. Nature Structural Biology. 1997; 4(3):215–222. https://doi.org/10.1038/nsb0397-215, doi: 10.1038/nsb0397-215.

Sonntag M, Jagtap PKA, Simon B, Appavou MS, Geerlof A, Stehle R, Gabel F, Hennig J, Sattler M. Segmental, Domain-Selective Perdeuteration and Small-Angle Neutron Scattering for Structural Analysis of Multi-Domain Proteins. Angewandte Chemie - International Edition. 2017; 56(32):9322–9325. doi: 10.1002/anie.201702904.

Souza PCT, Alessandri R, Barnoud J, Thallmair S, Faustino I, Grünewald F, Patmanidis I, Abdizadeh H, Bruininks BMH, Wassenaar TA, Kroon PC, Melcr J, Nieto V, Corradi V, Khan HM, Domański J, Javanainen M, Martinez-Seara H, Reuter N, Best RB, et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. Nature Methods. 2021; 18(4):382–388. doi: 10.1038/s41592-021-01098-3.

Stark AC, Andrews CT, Elcock AH. Toward optimized potential functions for protein-protein interactions in aqueous solutions: osmotic second virial coefficient calculations using the MARTINI coarse-grained force field. Journal of chemical theory and computation. 2013 9; 9(9). doi: 10.1021/ct400008p.

433 **Tesei G**, Martins JM, Kunze MBA, Wang Y, Crehuet R, Lindorff-Larsen K. DEER-PREdict: Software for efficient cal-
434 culation of spin-labeling EPR and NMR data from conformational ensembles. {PLOS} Computational Biology.
435 2021 1; 17(1):e1008551. https://doi.org/10.1371/journal.pcbi.1008551, doi: 10.1371/journal.pcbi.1008551.

436 **Tesei G**, Schulze TK, Crehuet R, Lindorff-Larsen K. Accurate model of liquid-liquid phase behaviour
437 of intrinsically-disordered proteins from optimization of single-chain properties. bioRxiv. 2021
438 1; p. 2021.06.23.449550. http://biorxiv.org/content/early/2021/09/10/2021.06.23.449550.abstract, doi:
439 10.1101/2021.06.23.449550.

440 **Tsanai M**, Frederix PWJM, Schroer CFE, Souza PCT, Marrink SJ. Coacervate formation studied by explicit solvent
441 coarse-grain molecular dynamics with the Martini model. Chem Sci. 2021; 12(24):8521–8530. http://dx.doi.
442 org/10.1039/D1SC00374G, doi: 10.1039/D1SC00374G.

443 **Vijay-Kumar S**, Bugg CE, Cook WJ. Structure of ubiquitin refined at 1.8Åresolution. Journal of Molecular
444 Biology. 1987; 194(3):531–544. https://www.sciencedirect.com/science/article/pii/0022283687906796, doi:
445 https://doi.org/10.1016/0022-2836(87)90679-6.

446 **Wang I**, Hennig J, Jagtap PKA, Sonntag M, Valcárcel J, Sattler M. Structure, dynamics and RNA binding of the
447 multi-domain splicing factor TIA-1. Nucleic Acids Research. 2014; 42(9):5949–5966. doi: 10.1093/nar/gku193.

448 **Wassenaar TA**, Ingólfsson HI, Böckmann RA, Tieleman DP, Marrink SJ. Computational lipidomics with insane:
449 A versatile tool for generating custom membranes for molecular simulations. Journal of Chemical Theory
450 and Computation. 2015; 11(5):2144–2155. doi: 10.1021/acs.jctc.5b00209.

451 **Wassenaar TA**, Pluhackova K, Böckmann RA, Marrink SJ, Tieleman DP. Going Backward: A Flexible Geometric
452 Approach to Reverse Transformation from Coarse Grained to Atomistic Models. Journal of Chemical Theory
453 and Computation. 2014 2; 10(2):676–690. https://doi.org/10.1021/ct400617g, doi: 10.1021/ct400617g.

454 **Wright PE**, Dyson HJ. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm.
455 Journal of Molecular Biology. 1999; 293(2):321–331. doi: 10.1006/jmbi.1999.3110.