Recalibration of protein interactions in Martini 3

F. Emil Thomasen[1†], **Tórur Skaalum**[1†], **Ashutosh Kumar**[2], **Sriraksha Srinivasan**[2], **Stefano Vanni**[2*], **Kresten Lindorff-Larsen**[1*]

***For correspondence:**
lindorff@bio.ku.dk (KLL);
stefano.vanni@unifr.ch (SV)

[†]These authors contributed equally to this work

[1]Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark; [2]Department of Biology, University of Fribourg, Fribourg, Switzerland

Abstract Multidomain proteins with flexible linkers and disordered regions play important roles in many cellular processes, but characterizing their conformational ensembles is difficult. In simulations, the situation is complicated further in multi-component systems—such as in the presence of a membrane—since the conformational ensemble depends on subtle balances between the interactions between and within protein, membrane, and water. We have previously shown that, for intrinsically disordered proteins (IDPs) and a small set of multidomain proteins, the widely used coarse grained force field, Martini 3, produces too compact ensembles in solution, and that increasing the strength of protein-water interactions in Martini 3 (by 10%) improves the agreement between simulations and small-angle X-ray scattering (SAXS) for these proteins. Here, we examine whether, as an alternative approach, decreasing the strength of interactions between protein beads can provide equivalent or further improved agreement with the experimental data, and explore the effects of these choices on the interactions with lipid bilayers. We have expanded the set of multidomain proteins to include a wider variety of sizes and domain architectures. Consistent with our previous results, we find that Martini 3 underestimates the global dimensions of this set of multidomain proteins, and that increasing the strength of protein-water interactions (by 10%) or decreasing the strength of non-bonded interactions between protein beads (by 12%) substantially improves the agreement with experimental SAXS data. We show that the 'symmetry' between rescaling protein-water and protein-protein interactions breaks down when studying interactions with membranes, and that rescaling protein-protein interactions better preserves the binding specificity of peripheral membrane proteins, multidomain proteins, and IDPs with lipid membranes. We conclude that decreasing the strength of protein-protein interactions improves the accuracy of Martini 3 for IDPs and multidomain proteins, both in solution and in the presence of a lipid membrane, providing a favorable alternative to rescaling protein-water interactions.

Introduction

Intrinsically disordered proteins (IDPs), folded proteins with long disordered tails, and multidomain proteins with folded domains connected by flexible linkers, are characterized by their high level of conformational dynamics. Molecular dynamics (MD) simulations provide a valuable tool for studying IDPs and multidomain proteins, as they can be used to determine full conformational ensembles at atomic resolution (***Thomasen and Lindorff-Larsen, 2022***). However, there are two central challenges that must be overcome for MD simulations to provide a useful description of such systems: the force field describing all the bonded and non-bonded interactions between atoms in the system must be sufficiently accurate *and* the conformational space of the protein must be

sufficiently sampled (***Bottaro and Lindorff-Larsen, 2018***).

One way to address the challenge of sufficient sampling is to use coarse-grained (CG) MD simulations in which groups of atoms are represented as single beads (***Ingólfsson et al., 2014***). Martini is a widely used CG model in which 2–4 non-hydrogen atoms are represented by a single bead (***Marrink et al., 2007***; ***Monticelli et al., 2008***). An attractive aspect of Martini is its modular structure and high degree of transferability, which allows the simulation of complex systems containing several different classes of biomolecules. The current version of Martini, Martini 3, shows improvements over previous versions in areas such as molecular packing, transmembrane helix interactions, protein aggregation, and DNA base pairing (***Souza et al., 2021***).

We have previously shown that Martini 3 simulations of IDPs produce overly compact conformational ensembles, resulting in poor agreement with small-angle X-ray scattering (SAXS) and paramagnetic relaxation enhancement (PRE) experiments (***Thomasen et al., 2022***). Using an approach inspired by previous work on assessing and rebalancing non-bonded interactions in Martini (***Stark et al., 2013***; ***Javanainen et al., 2017a***; ***Berg et al., 2018***; ***Berg and Peter, 2019***; ***Alessandri et al., 2019***; ***Larsen et al., 2020***; ***Benayad et al., 2021***; ***Majumder and Straub, 2021***; ***Lamprakis et al., 2021***; ***Martin et al., 2021***) and atomistic force fields (***Best et al., 2014***), we found that agreement with SAXS and PRE data could be significantly improved by uniformly increasing the strength of non-bonded interactions between protein and water beads by ~10% (***Thomasen et al., 2022***). This was also shown to be the case for three multidomain proteins, hnRNPA1, hisSUMO-hnRNPA1, and TIA1; however, due to the small sample size and the similarity between these three proteins, it remains an open question whether the approach generalizes to other multidomain proteins.

Our previous work was concerned with the properties of proteins in aqueous solution in the absence of other classes of biomolecules. Intuitively, increasing the strength of protein-water interactions should affect the affinity between proteins and other biomolecules. As a prototypical example, one would expect that increasing protein-water interactions would decrease the affinity of proteins for lipid membranes, since the interaction is tuned by the relative affinity of proteins for water vs. the membrane environment. The extent to which our previously described force field modification affects protein-membrane interactions, however, remains unclear. There is increasing evidence that IDPs and disordered regions play important physiological roles at lipid membranes (***Kjaergaard and Kragelund, 2017***; ***Zeno et al., 2018***; ***Das and Eliezer, 2019***; ***Fakhree et al., 2019***; ***Cornish et al., 2020***), and so it is important to understand better how force field changes that improve the description of disordered proteins in solution affect their interactions with membranes. In this context, it is important to note that unmodified Martini 3 has been quite successful at reproducing the specific membrane interactions for peripheral membrane proteins, as we previously showed (***Srinivasan et al., 2021***).

For previous versions of Martini, problems with overestimated protein-protein interactions have been corrected either by increasing the strength of protein-water interactions (***Berg et al., 2018***; ***Berg and Peter, 2019***; ***Larsen et al., 2020***; ***Martin et al., 2021***) or by decreasing the strength of interactions between protein beads (***Stark et al., 2013***; ***Javanainen et al., 2017a***; ***Benayad et al., 2021***). We hypothesize that for proteins in solution, the two force field corrections likely have similar effects, simply rebalancing the relative energies associated with hydration versus self-interaction. However, in the case of mixed systems, for example with proteins, water, and membranes, we might expect clearer differences between these approaches. For example, decreasing the strength of protein-protein interactions may better retain the affinity between proteins and other molecules as originally parameterized, while increased protein-water interactions may lower this affinity (Fig. 1). Thus, it remains an open question whether this specificity is retained when protein-water interactions are increased, and whether rescaling protein-protein interactions provides equivalent or improved agreement with experimental observations, both in comparison with unmodified Martini 3 and Martini 3 with rescaled protein-water interactions.

Here, we expand upon our previous work to address these questions. First, we have expanded the set of multidomain proteins to include 15 proteins for which SAXS data have previously been

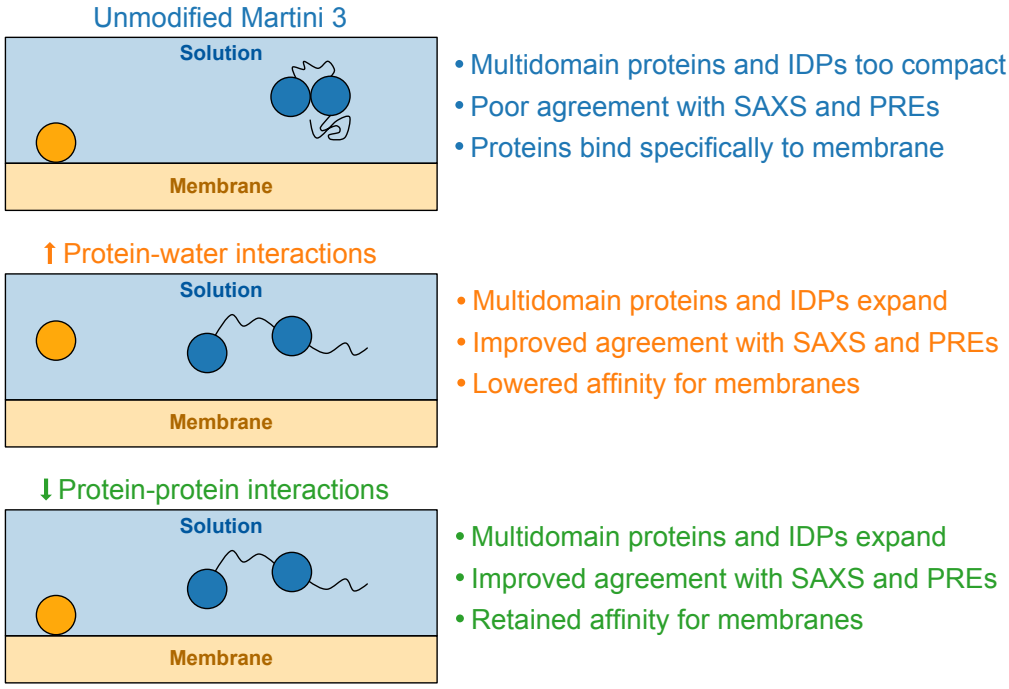

Figure 1. Expected effects of proposed force field modifications. Schematic overview showing the expected effects of rescaling protein-water and protein-protein interactions in Martini 3. Overestimated compactness of soluble IDPs and multidomain proteins and specific membrane interactions for peripheral membrane proteins have previously been reported (***Srinivasan et al., 2021***; ***Thomasen et al., 2022***).

collected (Fig. 2). Using this five-times larger set of proteins, we show that, as was the case for IDPs, increasing the strength of protein-water interactions by 10% improves the agreement with SAXS data. We further show that decreasing the strength of non-bonded interactions between protein beads by 12% leads to a comparable improvement in agreement with SAXS and PRE data for IDPs and multidomain proteins in solution, but better preserves the specificity of protein-membrane interactions for peripheral membrane proteins.

Results

Analysis of an expanded set of multidomain proteins

Previously, we tested Martini 3 using a set of three multidomain proteins, TIA1, hnRNPA1, and hisSUMO-hnRNPA1, for which SAXS data have been measured (***Sonntag et al., 2017***; ***Martin et al., 2021***). Given the similarity of the three proteins (all three are RNA-binding proteins, and the two latter differ only in the addition of a hisSUMO-tag), we wished to expand the set of proteins with mixed regions of order and disorder to include a wider range of sizes and domain architectures. We searched the literature for such proteins with reported SAXS data and identified 12 proteins that we added to our set (Fig. 2): the tri-helix bundle of the m-domain and the C2 domain of myosin-binding protein C (MyBP-$C_{MTHB-C2}$) (***Michie et al., 2016***); the C5, C6, and C7 domains of myosin-binding protein C (MyBP-$C_{C5-C6-C7}$) (***Nadvi et al., 2016***); linear di- tri- and tetraubiquitin (Ubq_2, Ubq_3, Ubq_4) (***Jussupow et al., 2023***); the two fluorescent proteins mTurquoise2 and mNeonGreen connected by a linker region with the insertion of 0, 8, 16, 24, 32, or 48 GS repeats (mTurq-GS_X-mNeon) (***Moses et al., 2022***); and Galectin-3 (Gal-3) (***Lin et al., 2017***). Apart from Gal-3, these proteins all contain at least two distinct folded domains, connected by linkers of different lengths and composition; three proteins (Gal-3, hnRNPA1, and hisSUMO-hnRNPA1) also contain a long disordered region attached to a folded domain. Collectively, we will refer to this set as multidomain proteins, though we note that Gal-3 only contains a single folded domain.

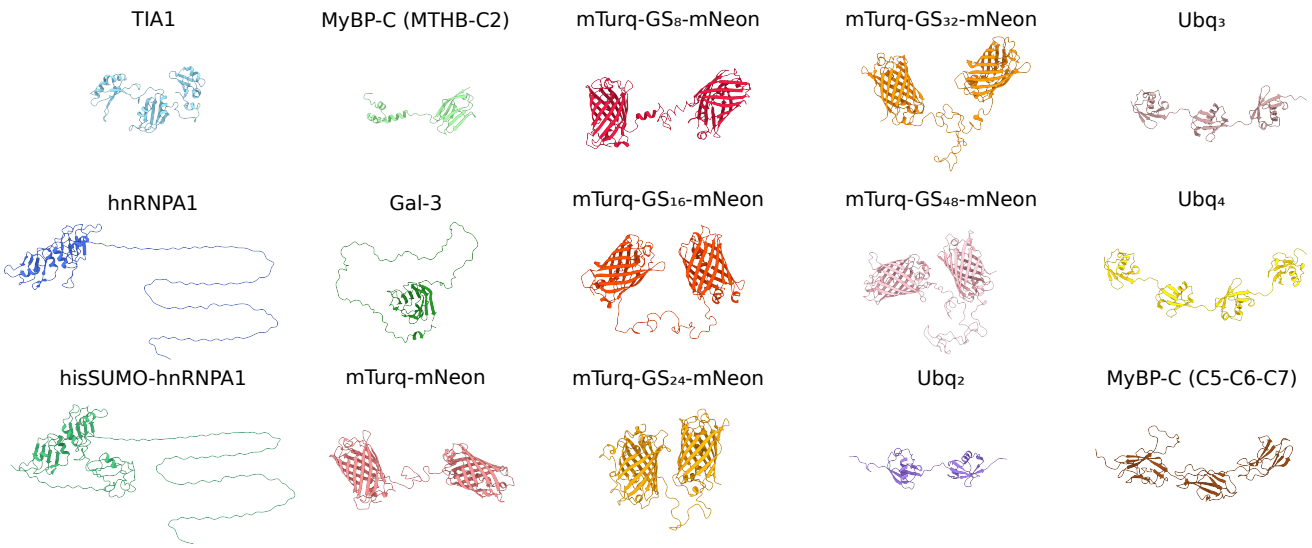

Figure 2. Starting structures for simulations of multidomain proteins. Starting structures of multidomain proteins used for Martini simulations. See the Methods section for a description of the source of the structures and how they were assembled.

We have previously shown that Martini 3 produces conformational ensembles that are more compact than found experimentally for a set of 12 IDPs and for the three multidomain proteins TIA1, hnRNPA1, and hisSUMO-hnRNPA1, and that rescaling ε in the Lennard-Jones potential between all protein and water beads by a factor λ_{PW}=1.10 resulted in more expanded ensembles that substantially improved the agreement with SAXS data (***Thomasen et al., 2022***). Using our much larger set of multidomain proteins, we examined whether Martini 3 generally produces too compact conformational ensembles of multidomain proteins, and whether our modified force field with rescaled protein-water interactions would generalize to the expanded set of proteins. We ran Martini 3 simulations of the 12 new multidomain proteins with unmodified Martini 3 and with λ_{PW}=1.10 and calculated SAXS intensities from the simulations. We found that, on average across the 15 proteins, increasing the strength of protein-water interactions by λ_{PW}=1.10 substantially improved the direct agreement with the experimental SAXS data, as quantified by the reduced χ^2, χ^2_r (Fig. 3). For only one of the 15 proteins, MyBP-$C_{MTHB-C2}$, the modified force field gave rise to reduced agreement with the SAXS data. This result shows that our previously proposed modification of protein-water interactions in Martini 3, which was optimized to improve the global dimensions of IDPs, also provides a general improvement in the global dimensions of multidomain proteins.

Rescaling protein-protein interactions

Inspired by previous work on earlier versions of the Martini force field, (***Stark et al., 2013***; ***Javanainen et al., 2017b***; ***Benayad et al., 2021***), we next examined whether rescaling protein-protein interactions instead of protein-water interactions would provide a similar or further improvement in the agreement with the experimental data. To do so, we ran Martini 3 simulations for the set of 12 IDPs with SAXS data available that we had studied previously (***Thomasen et al., 2022***) and the new set of 15 multidomain proteins. In these simulations, we rescaled ε in the Lennard-Jones potential between all protein beads by a factor λ_{PP}. We scanned different values of this parameter, and found λ_{PP}=0.88 to provide the best agreement with experiments (Fig. S1). We found that this level of rescaling protein-protein interactions (λ_{PP}=0.88) provided a comparable improvement in the agreement with the experimental data as rescaling protein-water interactions by λ_{PW}=1.10 for both multidomain proteins (Fig. 3) and IDPs (Fig. 4). To further test the effect of rescaling protein-protein interactions by λ_{PP}=0.88 and compare with the approach of rescaling protein-water interac-

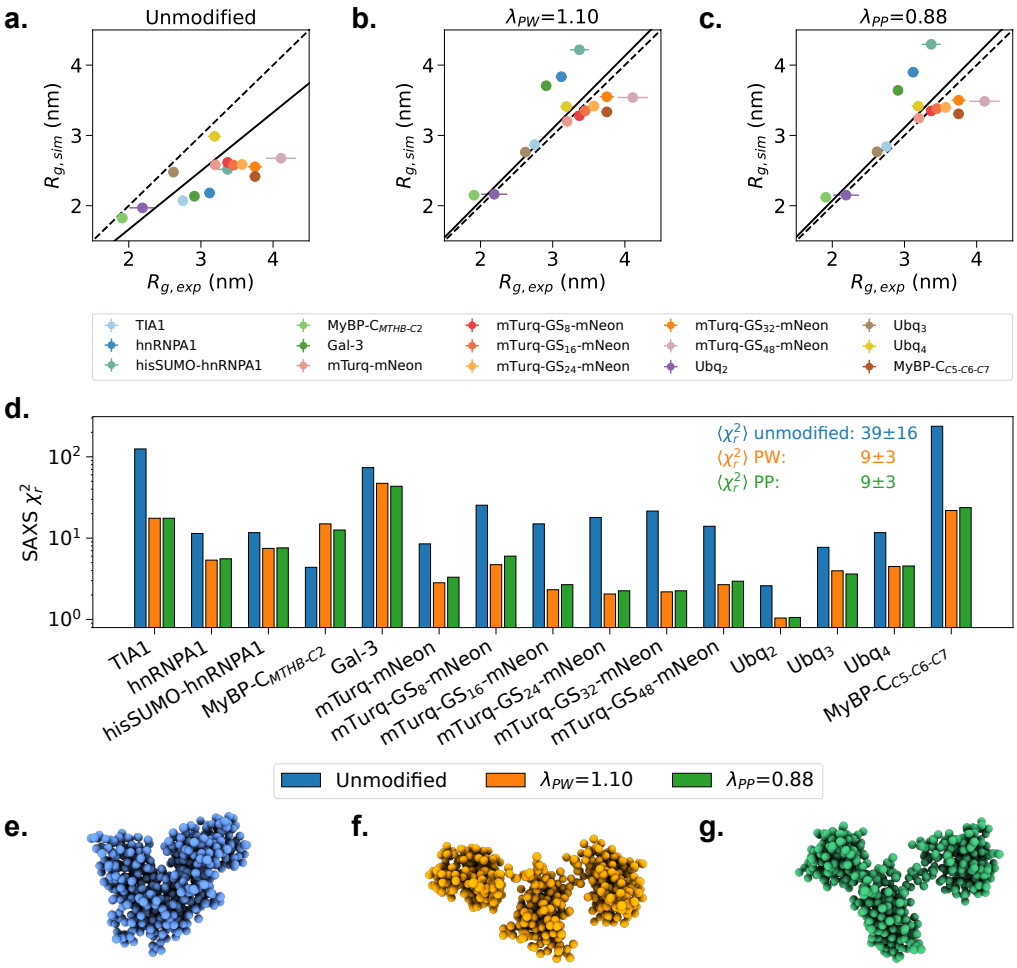

Figure 3. Agreement between simulations and SAXS data for multidomain proteins. R_g calculated from simulations plotted against R_g determined from Guinier fits to the SAXS data for **a** simulations with unmodified Martini 3, **b** simulations with protein-water interactions in Martini 3 rescaled by λ_{PW}=1.10, and **c** simulations with protein-protein interactions in Martini 3 rescaled by λ_{PP}=0.88. **d.** Reduced χ^2 to experimental SAXS intensities given by SAXS intensities calculated from unmodified Martini 3 simulations (blue) and Martini 3 simulations with protein-water interactions rescaled by λ_{PW}=1.10 (orange) or protein-protein interactions rescaled by λ_{PP}=0.88 (green). Mean and standard error of the mean over all proteins are shown on the plot. Note the logarithmic scale for χ_r^2. Representative conformation of TIA1 with an R_g corresponding to the average R_g in **e** simulations with unmodified Martini 3, **f** simulations with protein-water interactions in Martini 3 rescaled by λ_{PW}=1.10, and **g** simulations with protein-protein interactions in Martini 3 rescaled by λ_{PP}=0.88. Simulations of hnRNPA1, hisSUMO-hnRNPA1 and TIA1 with λ_{PW}=1.10 were taken from ***Thomasen et al.*** (***2022***).

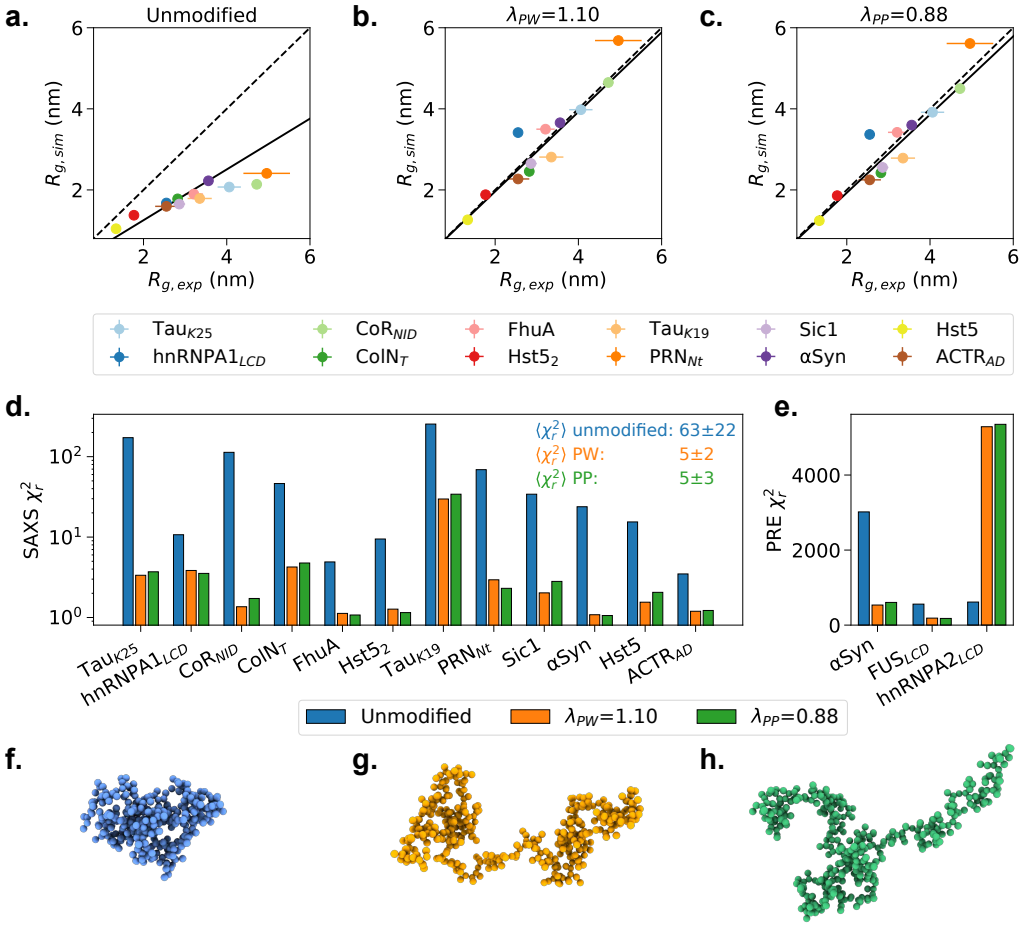

Figure 4. Agreement between simulations and SAXS or PRE data for IDPs. R_g calculated from simulations plotted against R_g determined from Guinier fits to the SAXS data for **a** simulations with unmodified Martini 3, **b** simulations with protein-water interactions in Martini 3 rescaled by λ_{PW}=1.10, and **c** simulations with protein-protein interactions in Martini 3 rescaled by λ_{PP}=0.88. **d.** Reduced χ^2 to experimental SAXS intensities given by SAXS intensities calculated from unmodified Martini 3 simulations (blue) and Martini 3 simulations with protein-water interactions rescaled by λ_{PW}=1.10 (orange) or protein-protein interactions rescaled by λ_{PP}=0.88 (green). Mean and standard error of the mean over all proteins are shown on the plot. Note the logarithmic scale for χ_r^2. Representative conformation of Tau$_{K25}$ with an R_g corresponding to the average R_g in **e** simulations with unmodified Martini 3, **f** simulations with protein-water interactions in Martini 3 rescaled by λ_{PW}=1.10, and **g** simulations with protein-protein interactions in Martini 3 rescaled by λ_{PP}=0.88. All simulations with λ_{PW}=1.10 were taken from ***Thomasen et al.*** (***2022***).

tions, we ran simulations of three IDPs with intramolecular PRE data available, the LCD of hnRNPA2 (***Ryan et al., 2018***), the LCD of FUS (***Monahan et al., 2017***), and α-synuclein (***Dedmon et al., 2005***), and calculated PRE data from the simulations (Fig. 4e). Again, λ_{PP}=0.88 provided the same level of agreement with the PRE data as we previously found using λ_{PW}=1.10 (***Thomasen et al., 2022***). Specifically, the agreement with the PRE data improved for α-synuclein and the FUS LCD, while the agreement worsened for the hnRNPA2 LCD (Fig. 4e). We concluded that decreasing the strength of protein-protein interactions by λ_{PP}=0.88 provides an equally good alternative to rescaling protein-water interactions for IDPs and multidomain proteins in solution.

While simulations of multidomain in solutions are affected by the interactions between the folded domains, we also tested more directly the effect of rescaling protein-protein interactions on the interactions between folded proteins. To this aim, we ran MD simulations of two protein systems that should undergo transient homodimerization, ubiquitin and villin HP36, which we

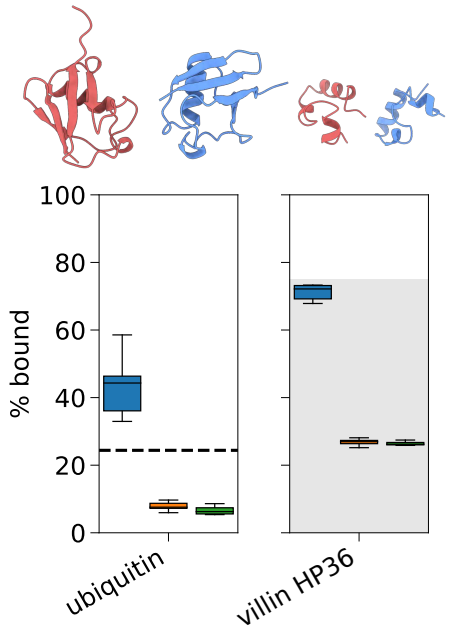

Figure 5. Protein-protein interactions. Fraction bound calculated from MD simulations of two copies of the folded proteins ubiquitin and villin HP36 with unmodified Martini 3 (blue) and Martini 3 with protein-water interactions rescaled by λ_{PW}=1.10 (orange) (taken from ***Thomasen et al.*** (***2022***)), and Martini 3 with protein-protein interactions rescaled by λ_{PP}=0.88 (green). Box plots show the results of 10 replica simulations. The bound fraction in agreement with Kd = 4.9 mM for ubiquitin self-association is shown as a dashed line (***Liu et al., 2012***). The bound fraction in agreement with a Kd > 1.5 mM for villin HP36 self-association is shown as a shaded gray area (***Brewer et al., 2005***).

also used in our previous work (***Thomasen et al., 2022***). Ubiquitin homodimerizes with a K_d of 4.9±0.3 mM based on NMR chemical shift perturbations (***Liu et al., 2012***) and villin HP36 should self-associate with a K_d > 1.5 mM based on NMR diffusion measurements (***Brewer et al., 2005***). We ran MD simulations of two copies of the proteins with λ_{PP}=0.88 and calculated the fraction of the time that the proteins were bound (Fig. 5). For both proteins λ_{PP}=0.88 resulted in decreased self-association, and again we found that λ_{PP}=0.88 gave comparable results to our previously published simulations with λ_{PW}=1.10 (***Thomasen et al., 2022***). Comparing the simulations with the expected fraction bound based on the experimentally determined K_d values, we found that ubiquitin self-association is likely slightly overestimated with unmodified Martini 3 and slightly underestimated with λ_{PP}=1.10 and λ_{PP}=0.88. For villin HP36, all three force fields gave rise to a fraction bound within the expected range. While the overestimated compaction of multidomain proteins suggest that interactions between folded domains may be too strong in Martini 3, our results on the self-association of ubiquitin and villin HP36 do not provide a clear indication that this is the case.

Rescaling protein-water interactions for backbone beads only

While the overall agreement with SAXS experiments was improved for almost all proteins when rescaling protein-protein or protein-water interactions, some proteins were still too expanded or compact with respect to the experimental R_g, suggesting that some sequence-specific effects on compaction were not fully captured. We reasoned that sequence-specific effects on the ensemble properties would possibly be better captured if we rescaled only the interactions between the protein backbone and water; this approach could lead to the desired expansion of the proteins while retaining the interactions of the amino acid side chains as originally parameterized. We therefore performed simulations of our set of IDPs and multidomain proteins in which we rescaled ε in the

Lennard-Jones potential between all protein backbone and water beads by a factor $\lambda_{\text{PW-BB}}$, scanning different values of this parameter, and found $\lambda_{\text{PW-BB}}$=1.22 to provide the best agreement with experiments (Fig. S2). However, the simulations of the IDPs and multidomain proteins with $\lambda_{\text{PW-BB}}$=1.22 showed similar agreement with experiments as when rescaling all protein-water interactions or protein-protein interactions (Figs. S3 and S4). Given that rescaling of only protein backbone-water interactions did not show any substantial improvement with respect to the previous approaches, and that the strong interactions between the protein backbone and water may have undesirable effects on the behavior of the hydration shell, we decided not to pursue this approach further.

Protein-membrane interactions

In the simulations described above, we found that the effects of increasing protein-water interactions or decreasing protein-protein interactions were very similar. We, however, hypothesized that these two force field modifications could have substantially different effects on systems in which proteins interact with other classes of molecules that are not protein or water. We expected that increased protein-water interactions would result in lower affinity for other molecules, which bind in competition with solvation, while decreased protein-protein interactions would not affect the affinity to the same extent, barring any effects of altering the conformational ensemble.

To examine the effect of rescaling the Lennard-Jones interaction parameters on the affinity of proteins for different biomolecules, we chose to investigate protein interactions with lipid membranes. We had two main motivations for this choice: first, protein-membrane interactions have been thoroughly characterized using Martini (***Yamamoto et al., 2015***; ***Naughton et al., 2016***; ***Srinivasan et al., 2021***); second, Martini has from its early development days in particular been focused on lipid membranes and protein-membrane interactions (***Marrink and Tieleman, 2013***; ***Herzog et al., 2016***; ***Javanainen et al., 2017a***).

We therefore performed simulations of peripheral proteins in the presence of lipid bilayers, using both unmodified Martini 3 and the two modified versions, λ_{PP}=0.88 and λ_{PW}=1.10, following a protocol we have previously described (***Srinivasan et al., 2021***). In short, we ran unbiased MD simulations starting with the protein at a minimum distance of 3 nm away from the bilayer. Over the course of the MD simulation, the proteins interact, often transiently and reversibly, with the membrane (Figs. S6 and S7), and membrane binding was quantified as previously described (***Srinivasan et al., 2021***) based on defining bound states when the minimum distance was lower than or equal to 0.7 nm.

To characterize the effect of our rescaling protocol on a broad set of protein-membrane interactions, we selected a diverse set of proteins: (i) one negative control, hen egg-white lysozyme, which is highly soluble in water and is not expected to interact specifically with the membrane (***Howard et al., 1988***); (ii) three peripheral membrane proteins consisting of a single folded domain (Phospholipase2, Arf1 in its GTP-bound state, and the C2 domain of Lactadherin) for which we previously characterized the membrane-binding behaviour (***Srinivasan et al., 2021***); (iii) two membrane-binding multidomain proteins: PTEN (1–351), containing a N-terminal Phosphatase domain and C2 domain that are known to be sufficient for membrane binding, and the Talin FERM domain, that has multiple sub-domains (F0 to F3) and binds to membranes through specific phosphoinositol(4,5)phosphate (PIP2) binding sites present in its F2 and F3 subdomains (***Buhr et al., 2023***); (iv) two intrinsically disordered regions (IDRs) that have been characterized as membrane-binding regions: the N-terminal IDR of TRPV4 (***Goretzki et al., 2022***) and a short C-terminal motif (CTM) of Complexin (***Snead et al., 2014***). For the two IDRs, simulations in solution with both λ_{PP}=0.88 and λ_{PW}=1.10 result in expanded ensembles and a larger average value of R_g compared to unmodified Martini 3 (Fig. S8).

As hypothesized, the different force field modifications have different effects on protein-membrane interactions (Fig. 6). In particular, we find that simulations with decreased protein-protein interactions (λ_{PW}=0.88) provide a similar degree of protein-membrane interaction when compared with unmodified Martini-3. In contrast, simulations with an increased strength of protein-water interac-

tions (λ_{PW}=1.10) show significantly reduced membrane affinity and binding for all proteins, almost always leading to a complete lack of interactions between the protein and the lipid bilayer. Importantly, the distinction between the membrane binding proteins and Lysozyme, which should not interact with the membrane, is retained with λ_{PP}=0.88, but not with λ_{PW}=1.10. Given that λ_{PW}=1.10 and λ_{PP}=0.88 provide a comparably good description of IDPs and multidomain proteins in solution, and that λ_{PP}=0.88 more accurately retains the specificity and strength of protein-membrane interactions as originally parameterized in Martini 3, we suggest that λ_{PP}=0.88 is overall a more robust and transferable modification to Martini 3.

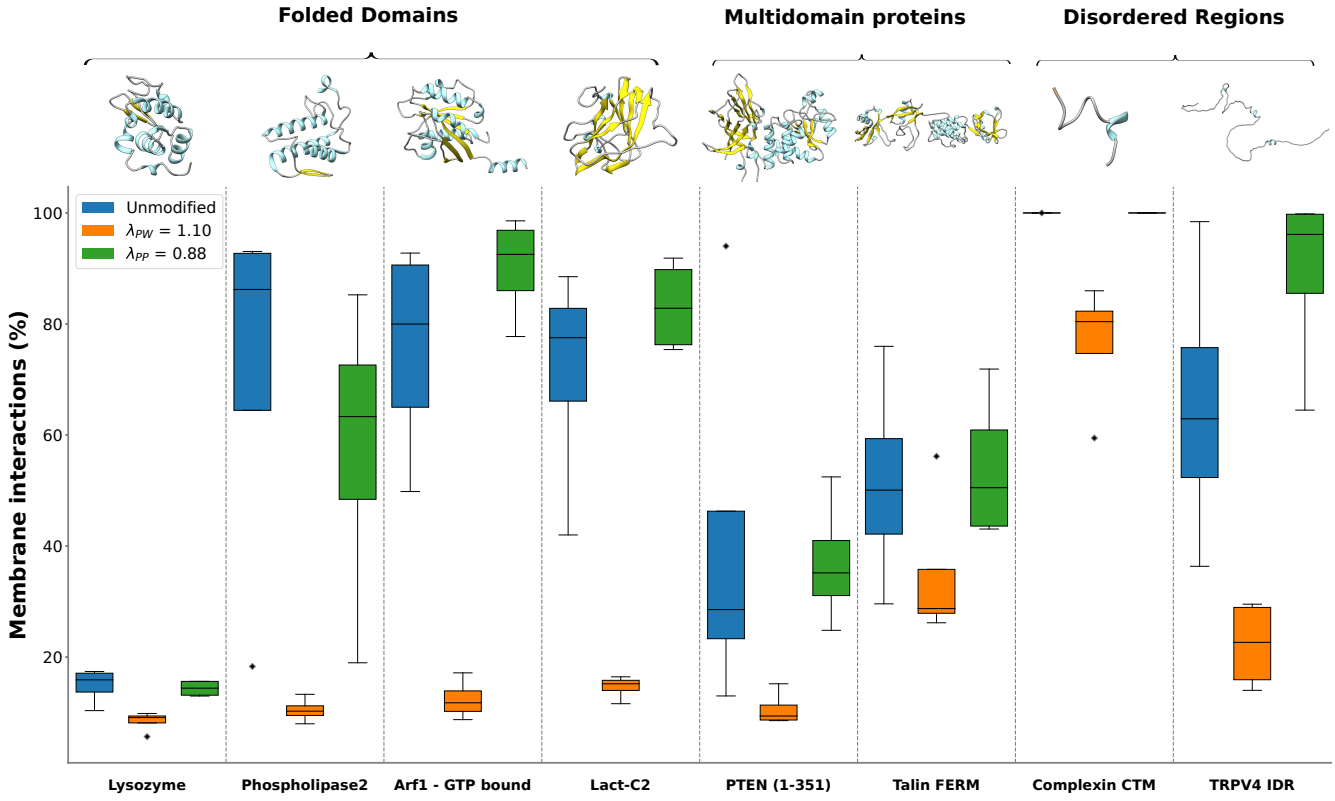

Figure 6. Protein-membrane interactions. MD simulations (four replicas, each 3 μs long) were performed for peripheral membrane proteins, multidomain proteins, and intrinsically disordered regions with appropriate membrane composition (see Methods for details). Simulations were performed with unmodified Martini 3 (blue), protein-water interactions in Martini 3 rescaled by λ_{PW}=1.10 (orange), and protein-protein interactions in Martini 3 rescaled by λ_{PP}=0.88 (green). For each system, the corresponding atomistic structure of the protein is shown on top.

Capturing effects of sequence changes

Having selected λ_{PP}=0.88 as the preferred force field modification, we next examined to what extent this force field could capture more subtle sequence effects in IDPs. We performed simulations of six variants of the LCD of hnRNPA1, which have varied composition of charged and aromatic residues while retaining the length of the wild-type sequence (***Bremer et al., 2022***), using unmodified Martini 3 and Martini 3 with λ_{PP}=0.88. We compared the R_g calculated from the simulations with R_g values measured by SAXS for the six variants and wild-type. As expected based on the results presented above, we found that unmodified Martini 3 substantially underestimated the R_g of all variants, while modifying protein-protein interactions by λ_{PP}=0.88 resulted in an expansion and slight overestimation of the R_g (Fig. 7 and S5). We found that unmodified Martini 3 did not accurately capture the variation in R_g associated with the sequence variation ($r_{Pearson}$=-0.08), while simulations with λ_{PP}=0.88 resulted in a more accurate estimate of the effect of the sequence variation on the R_g values ($r_{Pearson}$=0.88). This result demonstrates that Martini 3 with λ_{PP}=0.88 can, at

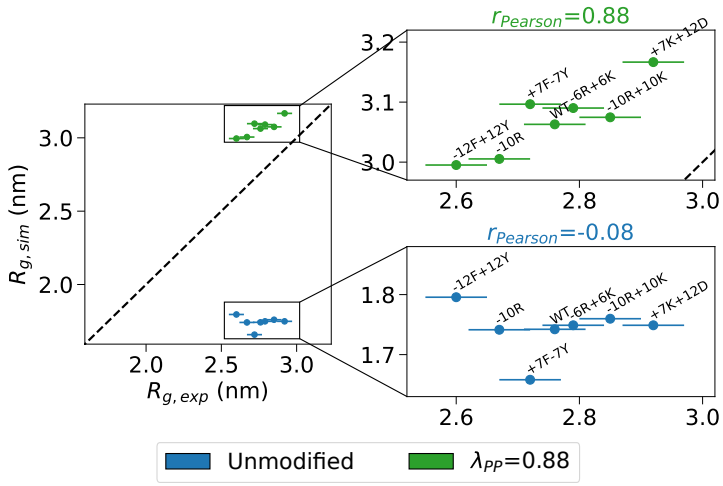

Figure 7. Radii of gyration of hnRNPA1$_{LCD}$ sequence variants. R_g calculated from simulations with unmodified Martini 3 (blue) and simulations with protein-protein interactions in Martini 3 rescaled by λ_{PP}=0.88 (green) are plotted against R_g determined by SAXS data for wild-type hnRNPA1$_{LCD}$ and six sequence variants with varied composition of charged and aromatic residues (***Bremer et al., 2022***). We show a zoom-in for each of the force fields along with the given Pearson correlation coefficient.

least in this case, capture the subtle effects of sequence variation on IDP ensembles.

Discussion

We have previously shown that simulations with Martini 3 underestimate the global dimensions of IDPs, and that increasing the strength of protein-water interactions by 10% results in more expanded ensembles and substantially improves the agreement with SAXS data (***Thomasen et al., 2022***). Here, we expanded this approach to a set of 15 multidomain proteins for which SAXS data have been recorded. Our results show that Martini 3 on average provides too compact ensembles of these multidomain proteins, and that, as was the case for IDPs, rescaling protein-water interactions by 10% substantially improves the agreement with SAXS data. We also show that decreasing the strength of interactions between protein beads by 12% results in the same expansion of the ensembles and improved agreement with experiments. We also tested the effect of increasing the strength of interactions between only the protein backbone beads and water, but did not find that this provides any further improvement in the agreement with the experimental data. While the different rescaling approaches provide essentially the same results for proteins in solution, we show that rescaling protein-protein interactions is the preferable option in order to best retain the specificity and strength of protein-membrane interactions as originally parameterized in Martini 3. An important outcome of our work is also the curation of a set of multidomain proteins with available SAXS data and starting structures for simulations, which can be used for future research in force field assessment and development.

One of the challenges when running Martini 3 simulations of multidomain proteins is selecting which regions to keep folded with the elastic network model and which regions to leave unrestrained. In this work, we manually selected the folded domains in the structures using domain annotations and intuition. It is, however, difficult to know *a priori* whether distinct domains should act as single structural modules due to specific interactions or move freely with respect to one another. Recently, it has been proposed to use the pairwise alignment error output from AlphaFold2 predictions to assign automatically the elastic network restraints (***Jussupow and Kaila, 2023***). In future work, this may provide a more accurate distinction between domains that should be relatively rigid or dynamic with respect to each other. Additionally, replacing the elastic network model with a more flexible structure-based model (***Go, 1983***) may provide the ability to sample both the

bound and unbound state in cases where folded domains have specific interactions (***Poma et al., 2017***).

Although the simple approach of decreasing the strength of protein-protein interactions uniformly by 12% shows an improvement over unmodified Martini 3 in reproducing the global dimensions of IDPs and multidomain proteins, we note that the agreement with the SAXS data is still not perfect (χ_r^2>1 in most cases), and there are systematic outliers with respect to the experimental R_g values. Although some of the system-specific deviations could potentially be alleviated by e.g. more accurately assigning and modeling the restraints on the folded domains, the overall deviation from the experimental data suggests that a more fundamental rebalancing of non-bonded interactions, and perhaps also CG mapping scheme, is necessary to describe the behavior of IDPs and multidomain proteins within the Martini framework. Again, we suggest that the data we have collected here will be useful to test any such changes, and the results obtained with λ_{PP}=0.88 are a useful point of reference for other force field modifications.

For other types of systems, it has been suggested that the non-bonded interactions in Martini 3 must be rescaled to a different extent to reach agreement with experimental observations. For example, modifying protein-water interactions in Martini 3 affects the propensity of the disordered LCD of FUS to form condensates in a way that appears to depend on the salt concentration (***Zerze, 2023***), while the insertion of transmembrane helices into the phospholipid bilayer may require decreased protein-water interactions (***Claveras Cabezudo et al., 2023***). Taken together with the results presented here, it seems that rescaling non-bonded interactions may not provide a universally transferable protein model in Martini.

Overall, however, our results demonstrate that for soluble proteins decreasing the non-bonded interactions between all protein beads by 12% leads to a more accurate balance of interactions while retaining the specificity of protein-membrane interactions. We foresee that our protocol will be a useful starting point to investigate the interactions of IDPs with lipid membranes using chemically transferable MD simulations, and that these investigations will further provide insights into possible strategies on future force field development efforts. Since CG simulations also play an important role in integrative structural biology (***Thomasen and Lindorff-Larsen, 2022***), we also expect that these developments will enable an even tighter link between simulations and experiments to study large and complex biomolecular assemblies.

Methods

IDP simulations

We performed MD simulations of a set of 12 IDPs with SAXS data available (Table 1) and three IDPs with intramolceular PRE data available (Table 2) (***Tesei et al., 2021b***; ***Thomasen et al., 2022***) using Gromacs 2020.3 (***Abraham et al., 2015***). We ran simulations with the Martini 3.0 force field (***Souza et al., 2021***) with the well-depth, ε, in the Lennard-Jones potential between all protein beads rescaled by a factor λ_{PP} or with ε in the Lennard-Jones potential between all protein backbone and water beads rescaled by a factor $\lambda_{PW\text{-}BB}$. We generated CG structures using Martinize2 based on initial all-atom structures corresponding to the 95th percentile of the R_g-distributions from simulations in ***Tesei et al.*** (***2021b***). Secondary structure and elastic restraints were not assigned for IDPs. Structures were placed in a dodecahedral box using Gromacs editconf and solvated, with NaCl concentrations corresponding to the ionic strength used in SAXS or PRE experiments, using the Insane python script (***Wassenaar et al., 2015***). The systems were equilibrated for 10 ns with a 2 fs time step using the Velocity-Rescaling thermostat (***Bussi et al., 2007***) and Parinello-Rahman barostat (***Parrinello and Rahman, 1981***). Production simulations were run for 40 µs with a 20 fs time step using the Velocity-Rescaling thermostat (***Bussi et al., 2007***) and Parinello-Rahman barostat (***Parrinello and Rahman, 1981***). The simulation temperature was set to match the SAXS or PRE experiment, and the pressure was set to 1 bar. Non-bonded interactions were treated with the Verlet cutoff scheme. A cut-off of 1.1 nm was used for van der Waals interactions. A dielectric

Protein	N_R	SAXS R_g (nm)	T (K)	c_s (M)	SAXS ref.
Hst5	24	1.34 ± 0.05	293	0.15	***Jephthah et al.* (*2019*)**
$Hst5_2$	48	1.77 ± 0.049	298	0.15	***Fagerberg et al.* (*2020*)**
$ACTR_{AD}$	71	2.55 ± 0.27	278	0.2	***Kjaergaard et al.* (*2010*)**
Sic1	92	2.86 ± 0.14	293	0.2	***Gomes et al.* (*2020*)**
$ColN_T$	98	2.82 ± 0.034	277	0.4	***Johnson et al.* (*2017*)**
Tau_{K19}	99	3.35 ± 0.29	288	0.15	***Mylonas et al.* (*2008*)**
$hnRNPA1_{LCD}$	137	2.55 ± 0.1	296	0.05	***Martin et al.* (*2020*)**
αSyn	140	3.56 ± 0.036	293	0.2	***Ahmed et al.* (*2021*)**
FhuA	144	3.21 ± 0.22	298	0.15	***Riback et al.* (*2017*)**
Tau_{K25}	185	4.06 ± 0.28	288	0.15	***Mylonas et al.* (*2008*)**
CoR_{NID}	271	4.72 ± 0.12	293	0.2	***Cordeiro et al.* (*2019*)**
PRN_{Nt}	334	4.96 ± 0.56	298	0.15	***Riback et al.* (*2017*)**

Table 1. IDPs with available SAXS data. Number of amino acid residues (N_R), experimental R_g, temperature (T), and salt concentration (c_s) used in simulations, and the reference for the SAXS data used.

Protein	N_R	T (K)	c_s (M)	PRE ref.
αSyn	140	283	0.125	***Dedmon et al.* (*2005*)**
$hnRNPA2_{LCD}$	155	298	0.005	***Ryan et al.* (*2018*)**
FUS_{LCD}	163	298	0.15	***Monahan et al.* (*2017*)**

Table 2. IDPs with available PRE data. Number of amino acid residues (N_R), temperature (T), and salt concentration (c_s) used in simulations, and the reference for the PRE data used.

constant of 15 and cut-off of 1.1 nm were used for Coulomb interactions. Simulation frames were saved every 1 ns. Molecule breaks from crossing the periodic boundaries were treated with Gromacs trjconv using the flags: -pbc whole -center. Convergence of the simulations was assessed by block-error analysis (***Flyvbjerg and Petersen, 1989***) of R_g calculated from simulation coordinates using the blocking code from: https://github.com/fpesceKU/BLOCKING. All CG trajectories were back-mapped to all-atom structures using a simplified version (***Larsen et al., 2020***) of the Backward algorithm (***Wassenaar et al., 2014***), in which simulation runs are excluded and the two energy minimization runs are shortened to 200 steps.

Multidomain protein structures

We performed MD simulations of a set of 15 multidomain proteins with SAXS data available (Table 3). We built the initial structure of MyBP-$C_{MTHB-C2}$ based on the NMR structure containing both domains (PDB: 5K6P) (***Michie et al., 2016***). We built the structures of the linear polyubiquitin chains, Ubq_2, Ubq_3, and Ubq_4, based on the crystal structure of the open conformation of Ubq_2 (PDB: 2W9N) (***Komander et al., 2009***). For Ubq_3 and Ubq_4, the linker regions between the original and extended structures were remodelled using Modeller (***Šali and Blundell, 1993***). We built the initial structure of Gal-3 based on the crystal structure of the folded C-terminal domain (PDB: 2NMO) (***Collins et al., 2007***) and the IDR from the AlphaFold structure of full-length Gal3 (AF-P17931-F1) (***Jumper et al., 2021***; ***Tunyasuvunakool et al., 2021***). We built the structure of MyBP-$C_{C5-C6-C7}$ based on the NMR structure of the C5 domain (PDB: 1GXE) (***Idowu et al., 2003***), and the AlphaFold structure of the full-length MyBP-C (AF-Q14896-F1) (***Jumper et al., 2021***; ***Tunyasuvunakool et al., 2021***). We inserted missing residues in the NMR structure of the C5 domain using Modeller (***Šali and Blundell, 1993***). For the mTurq-GS_X-mNeon constructs, we used structures from Monte-Carlo simulations in ***Moses et al.* (*2022*)** as starting structures for our simulations. To validate the starting structures, we calculated the RMSD between the two fluorescent protein domains and corresponding crystal structures (mTurquoise2 (PDB: 4AR7) (***von Stetten et al., 2012***) and mNeonGreen (PDB:

Protein	N_R	SAXS R_g (nm)	T (K)	c_s (M)	SAXS ref.
MyBP-$C_{MTHB-C2}$	137	1.91 ± 0.08	277	0.15	***Michie et al.* (*2016*)**
Ubq_2	162	2.2 ± 0.18	293	0.33	***Jussupow et al.* (*2023*)**
Ubq_3	228	2.62 ± 0.02	293	0.33	***Jussupow et al.* (*2023*)**
Gal-3	250	2.91 ± 0.06	303	0.04	***Lin et al.* (*2017*)**
TIA1	275	2.75 ± 0.05	300	0.1	***Sonntag et al.* (*2017*)**
Ubq_4	304	3.19 ± 0.09	293	0.33	***Jussupow et al.* (*2023*)**
hnRNPA1	314	3.12 ± 0.08	300	0.15	***Martin et al.* (*2021*)**
MyBP-$C_{C5-C6-C7}$	328	3.75 ± 0.08	298	0.28	***Nadvi et al.* (*2016*)**
hisSUMO-hnRNPA1	433	3.4 ± 0.13	300	0.1	***Martin et al.* (*2021*)**
mTurq-mNeon	470	3.20 ± 0.04	293	0.15	***Moses et al.* (*2022*)**
mTurq-GS_8-mNeon	486	3.37 ± 0.04	293	0.15	***Moses et al.* (*2022*)**
mTurq-GS_{16}-mNeon	502	3.45 ± 0.06	293	0.15	***Moses et al.* (*2022*)**
mTurq-GS_{24}-mNeon	518	3.57 ± 0.08	293	0.15	***Moses et al.* (*2022*)**
mTurq-GS_{32}-mNeon	534	3.8 ± 0.1	293	0.15	***Moses et al.* (*2022*)**
mTurq-GS_{48}-mNeon	566	4.1 ± 0.21	293	0.15	***Moses et al.* (*2022*)**

Table 3. Multidomain proteins with available SAXS data. Number of amino acid residues (N_R), experimental R_g, temperature (T), and salt concentration (c_s) used in simulations, and the reference for the SAXS data used.

5LTR) (***Clavel et al., 2016***)) using PyMOL align, which gave an RMSD of 0.2-0.3 Å.

Multidomain protein simulations

We ran MD simulations of the set of multidomain proteins using Gromacs 2020.3 (***Abraham et al., 2015***). We ran simulations with the Martini 3.0 force field (***Souza et al., 2021***), as well as several modified versions of Martini 3.0 in which the well-depth, ε, in the Lennard-Jones potential between all protein and water beads was rescaled by a factor λ_{PW}, ε in the Lennard-Jones potential between all protein beads was rescaled by a factor λ_{PP}, or ε in the Lennard-Jones potential between all protein backbone and water beads was rescaled by a factor $\lambda_{PW\text{-}BB}$. We assigned secondary structure-specific potentials using DSSP (***Kabsch and Sander, 1983***) and Martinize2. The secondary structure of all residues in linkers and IDRs were manually assigned to coil, turn, or bend. We applied an elastic network model using Martinize2 consisting of harmonic potentials with a force constant of 700 kJ mol^{-1} nm^{-2} between all backbone beads within a cut-off distance of 0.9 nm. We removed the elastic network potentials in all linkers and IDRs and between folded domains, so only the structures of individual folded domains were restrained (Table S1). Dihedral and angle potentials between sidechain and backbone beads were assigned using the -scfix flag in Martinize2, but removed in all linkers and IDRs. Structures were placed in a dodecahedral box using Gromacs editconf and solvated, with NaCl concentrations corresponding to the ionic strength used in SAXS experiments, using the Insane python script (***Wassenaar et al., 2015***). The systems were equilibrated for 10 ns with a 2 fs time step using the Berendsen thermostat and Berendsen barostat (***Berendsen et al., 1984***). Production simulations were run for 40 µs with a 20 fs time step using the Velocity-Rescaling thermostat (***Bussi et al., 2007***) and Parinello-Rahman barostat (***Parrinello and Rahman, 1981***). The simulation temperature was set to match the corresponding SAXS experiment and the pressure was set to 1 bar. Non-bonded interactions were treated with the Verlet cutoff scheme. A cut-off of 1.1 nm was used for van der Waals interactions. A dielectric constant of 15 and cut-off of 1.1 nm were used for Coulomb interactions. Simulation frames were saved every 1 ns. Molecule breaks from crossing the periodic boundaries were treated with Gromacs trjconv using the flags: -pbc whole -center. Convergence of the simulations was assessed by block-error analysis (***Flyvbjerg and Petersen, 1989***) of R_g calculated from simulation coordinates using the blocking code from: https://github.com/fpesceKU/BLOCKING. All CG trajectories were back-mapped to all-atom struc-

Protein	N_R	K_d (mM)	T (K)	c_s (M)	c_p (mM)	Self-association ref.
Villin HP36	36	>1.5	298	0.15	8.5	***Brewer et al.* (*2005*)**
Ubq	76	4.9 ± 0.3	303	0.11	1.0	***Liu et al.* (*2012*)**

Table 4. Proteins with available self-association data. Number of amino acid residues (N_R), experimental K_d for self-association, temperature (T), salt concentration (c_s), and initial protein concentration (c_p) used in simulations, and the reference for the self-association data.

tures using a simplified version (***Larsen et al., 2020***) of the Backward algorithm (***Wassenaar et al., 2014***), in which simulation runs are excluded and the two energy minimization runs are shortened to 200 steps.

Simulations of the self-association of folded proteins

We ran MD simulations of two copies of ubiquitin and two copies of villin HP36 as previously described (***Thomasen et al., 2022***). We used PDB ID 1UBQ (***Vijay-Kumar et al., 1987***) and PDB ID 1VII (***McKnight et al., 1997***) as starting structures for ubiquitin and villin HP36, respectively. The simulations 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 cubic box with side length 14.92 nm, giving a protein concentration of 1 mM. Two copies of villin HP36 were placed in a cubic box with side length 7.31 nm, giving a protein concentration of 8.5 mM. (ii) Secondary structure was assigned with DSSP (***Kabsch and Sander, 1983***) in Martinize2. (iii) An elastic network model was applied with Martinize2. The elastic restraints consisted of a harmonic potential of 700 kJ $mol^{-1}nm^{-2}$ between backbone beads within a 0.9 nm cutoff. For ubiquitin, we removed elastic restraints from the C-terminus (residues 72–76) to allow for flexibility (***Lindorff-Larsen et al., 2005***). (iv) Dihedral and angular potentials between side chains and backbone beads were added based on the initial structures with the -scfix flag in Martinize2. Simulations of ubiquitin were run with 110 mM NaCl at 303 K and simulations of villin HP36 were run with 150 mM NaCl at 298 K. For each protein, we ran 10 replica simulations of 40 µs per replica using Martini 3 modified with λ_{PP}=0.88.

We analyzed the population of the bound states in our simulations by calculating the minimum distance between beads in the two protein copies over the trajectory with Gromacs mindist. The fraction bound was defined as the fraction of frames where the minimum distance was below 0.8 nm. We calculated the expected fraction of bound protein at the concentrations in our simulations based on the K_d-values of 4.9 mM and 1.5 mM determined for ubiquitin and villin HP36 self-association respectively (***Liu et al., 2012***; ***Brewer et al., 2005***). The bound fraction was calculated as

$$\phi_b = \frac{4C_p + K_d - \sqrt{8K_dC_p + {K_d}^2}}{4C_p} \quad (1)$$

where ϕ_b is the bound fraction, C_p is the concentration of protein in the simulation box (using the average box volume over all simulation trajectories), and K_d is the dissociation constant.

hnRNPA1 LCD variant simulations

We ran MD simulations of a set of six variants of the hnRNPA1 LCD (-10R, -10R+10K, -12F+12Y, -6R+6K, +7F-7Y, +7K+12D) for which the R_g has previously been determined by SAXS experiments (***Bremer et al., 2022***). The variants contain substitutions to and from charged and aromatic residues, but have the same sequence length as the wild-type protein, and were selected to have a relatively large deviation in R_g from the wild-type; protein sequences can be found in the supporting information of ***Bremer et al.* (*2022*)**. We ran MD simulations with unmodified Martini 3 and Martini 3 with ε in the Lennard-Jones potential between all protein beads rescaled by a factor λ_{PP}=0.88. Simulations were set up using the same protocol as for the other IDPs described above. The systems

Protein	PDB ID	N_R	Bilayer composition
Lysozyme	1AKI	129	DOPC
Phospholipase2	1POA	118	DOPC
Arf1-GTP bound	2KSQ	181	DOPC
Lact-C2	3BN6	158	DOPC
PTEN (1-351)	AF-F6KD01-F1	351	DOPC:DOPS (8:2)
Talin's FERM	3IVF	368	POPC:PIP_2 (10% PIP_2 in upper leaflet)
Complexin CTM	Modelled with ColabFold	16	POPC:POPS (7:3)
TRPV4 IDR	AF-A0A1D5PXA5-F1	133	POPC:DOPS:PIP_2 (7:2:1)

Table 5. Membrane-protein systems. Structure (PDB ID), number of amino acid residues (N_R) in protein, and lipid composition in the membrane bilayer used in the simulations. Structures starting with AF prefix are AlphaFold-predicted structures (***Varadi et al., 2022***).

were equilibrated for 10 ns with a 2 fs time step using the Berendsen thermostat and Berendsen barostat (***Berendsen et al., 1984***). Production simulations were run for 100 µs with a 20 fs time step using the Velocity-Rescaling thermostat (***Bussi et al., 2007***) and Parinello-Rahman barostat (***Parrinello and Rahman, 1981***). Simulations were run with 150 mM NaCl at 298 K and 1 bar.

Peripheral membrane protein simulations

We performed MD simulations of one negative control, three peripheral membrane proteins, two multidomain proteins, and two intrinsically disordered regions with lipid bilayers of different compositions (Table 5). We ran simulations with the Martini 3 force field (***Souza et al., 2021***), or with modified force fields in which ε in (i) the Lennard-Jones potential between all protein beads were rescaled by a factor λ_{PP}=0.88 or (ii) with ε in the Lennard-Jones potential between all protein and water beads rescaled by a factor λ_{PW}=1.10.

Initial structures of proteins were obtained either from the RCSB database (***Rose et al., 2012***) or from the AlphaFold protein structure database (***Varadi et al., 2022***). For Complexin CTM, we used ColabFold v1.5.2 (***Mirdita et al., 2022***) to model 16-residues long (ATGAFETVKGFFPFGK) disordered region. The N-terminal IDR of TRPV4 (residues 2–134) was taken from the full-length AlphaFold structure of TRPV4 (A0A1D5PXA5). Initial structure of the FERM domains in Talin (PDB:3IVF) had missing residues (134–172), which we modelled using MODELLER (***Webb and Sali, 2016***) via the Chimera interface (***Pettersen et al., 2004***). CG structures of proteins were generated using Martinize2, with DSSP (***Kabsch and Sander, 1983***) flag to assign secondary structures. An elastic network was applied with a harmonic potential of a force constant 700 kJ mol^{-1} nm^{-2} between all backbone beads within a cutoff of 0.8 nm. We removed elastic network potentials between different domains and in linkers and in IDRs of multidomain proteins. Secondary structure and elastic network was not assigned to the two IDRs.

All the lipid bilayers, with initial lateral dimension of 20 nm × 20 nm, were generated using CHARMM-GUI Martini maker (***Qi et al., 2015***), except in the systems where phosphoinositol-(4,5)-phosphate (PIP2) lipids were needed, which instead were generated using the Insane python script (***Wassenaar et al., 2015***). We used the parameter for SAP2_45 lipids (***Borges-Araújo et al., 2021***) to model PIP2 in the bilayer. The bilayers generated from CHARMM-GUI were then minimized and equilibrated following the 6-step equilibration protocol. To compute protein-membrane interactions, systems were generated as previously described (***Srinivasan et al., 2021***), with a minimum distance of 3 nm between any bead of protein and any beads of lipid. Systems were first energy minimized using steepest descent algorithm after which a short MD run of 200 ps was performed with the protein backbone beads restrained. Production simulations (four replicas for each system) were run for 3 µs with a time step of 20 fs using velocity-rescale thermostat (***Bussi et al., 2007***) and Parrinello-Rahman barostat (***Parrinello and Rahman, 1981***).

We performed MD simulation of the two IDRs (Complexin CTM and TRPV4 IDR) in solution with

unmodified Martini 3 and both of the modified versions of Martini 3. For these simulations, we took the CG structure and placed it in a cubic box using Gromacs editconf, and solvated and ionized with a concentration of 150 mM of NaCl. Then the system was minimized for 10000 steps with steepest descent algorithm and a short equilibration run was performed with Berendsen thermostat and Berendsen barostat (***Berendsen et al., 1984***) with a time step of 2 fs. Production simulations were run for 10 µs with a 20 fs time-step using Parrinello-Rahman barostat (***Parrinello and Rahman, 1981***) and velocity-rescaling thermostat (***Bussi et al., 2007***). All the simulations were performed with GROMACS 2021.5 (***Abraham et al., 2015***). Initial 100 ns of production run were discarded from all the trajectories for further analysis.

SAXS calculations

We extracted 15,000 evenly distributed frames from each back-mapped trajectory to calculate SAXS profiles using Pepsi-SAXS (***Grudinin et al., 2017***). To avoid overfitting the parameters for the contrast of the hydration layer ($\delta\rho$) and the displaced solvent ($r0$) by fitting them individually to each structure, we used the fixed values for these parameters determined in ***Pesce and Lindorff-Larsen*** (***2021***). We globally fitted the scale and constant background with least-squares regression weighted by the experimental errors using Scikit-learn (***Pedregosa et al., 2011***). To assess the agreement between the experimental SAXS profiles and those calculated from simulations, we calculated the χ_r^2 between the ensemble-averaged calculated SAXS intensities (I_{calc}) and the experimental SAXS intensity (I_{exp}):

$$\chi_r^2 = \frac{1}{m}\sum_q^m \left(\frac{I_q^{calc} - I_q^{exp}}{\sigma_q^{exp}} \right)^2 \tag{2}$$

where σ^{exp} is the error of the experimental SAXS intensity and m is the number of measured SAXS intensities. We used the Bayesian Indirect Fourier Transform algorithm (BIFT) to rescale the errors of the experimental SAXS intensities, in order to obtain a more consistent error estimate across the different proteins (***Hansen, 2000***; ***Larsen and Pedersen, 2021***).

PRE calulcations

We used the DEER-PREdict software (***Tesei et al., 2021a***) to calculate PRE ratios from the back-mapped trajectories of α-synuclein, FUS_{LCD}, and $hnRNPA2_{LCD}$ (Table 2). DEER-PREDICT uses a rotamer library approach to model the MTSL spin-label (***Polyhach et al., 2011***) and a model-free formalism to calculate the spectral density (***Iwahara et al., 2004***). We assumed an effective correlation time of the spin label, τ_t, of 100 ps, a molecular correlation time, τ_c, of 4 ns (***Gillespie and Shortle, 1997***), a transverse relaxation rate for the diamagnetic protein of 10 s^{-1} and a total INEPT time of the HSQC measurement of 10 ms (***Battiste and Wagner, 2000***). The agreement between calculated and experimental PREs was assessed by calculating the χ_r^2 over all spin-label positions,

$$\chi_r^2 = \frac{1}{N_{labels}N_{res}}\sum_j^{N_{labels}}\sum_i^{N_{res}} \left(\frac{Y_{ij}^{exp} - Y_{ij}^{calc}}{\sigma_{ij}^{exp}} \right)^2 \tag{3}$$

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.

Radii of gyration

We calculated the R_g from CG simulation trajectories using Gromacs gyrate (***Abraham et al., 2015***) and calculated the error of the average R_g using block-error analysis (***Flyvbjerg and Petersen, 1989***) (https://github.com/fpesceKU/BLOCKING). Experimental R_g-values and corresponding error bars were calculated from SAXS profiles by Guinier analysis using ATSAS AUTORG with default settings

(***Petoukhov et al., 2007***), except in the case of the hnRNPA1$_{LCD}$ variants, for which we used the R_g-values reported in ***Bremer et al.*** (***2022***), which were determined from SAXS data using an empirical molecular form factor approach.

Data and software availability

Scripts and data will be available via https://github.com/KULL-Centre/_2023_Thomasen_Martini

Acknowledgments

We acknowledge the use of computational resources from Computerome 2.0 and the core facility for biocomputing at the Department of Biology. This research was supported by the Lundbeck Foundation BRAINSTRUC initiative (R155-2015-2666 to K.L.-L.) and the PRISM (Protein Interactions and Stability in Medicine and Genomics) centre funded by the Novo Nordisk Foundation (NNF18OC0033950, to K.L.-L.). SV and AK acknowledge support by the Swiss National Science Foundation through the National Center of Competence in Research Bio-Inspired Materials. This work was supported by grants from the Swiss National Supercomputing Centre (CSCS) under project ID s1176.

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 PCT, Thallmair S, Melo MN, de Vries AH, Marrink SJ. Pitfalls of the Martini Model. Journal of Chemical Theory and Computation. 2019 oct; 15(10):5448–5460. https://doi.org/10.1021/acs.jctc.9b00473, doi: 10.1021/acs.jctc.9b00473.

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.

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.

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics. 1984 oct; 81(8):3684–3690. https://doi.org/10.1063/1.448118, doi: 10.1063/1.448118.

Berg A, Kukharenko O, Scheffner M, Peter C. Towards a molecular basis of ubiquitin signaling: A dual-scale simulation study of ubiquitin dimers. PLoS Computational Biology. 2018; 14(11):1–14. doi: 10.1371/journal.pcbi.1006589.

Berg A, Peter C. Simulating and analysing configurational landscapes of protein-protein contact formation. Interface Focus. 2019; 9(3):20180062. https://royalsocietypublishing.org/doi/abs/10.1098/rsfs.2018.0062, doi: 10.1098/rsfs.2018.0062.

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.

Borges-Araújo L, Souza PC, Fernandes F, Melo MN. Improved parameterization of phosphatidylinositide lipid headgroups for the Martini 3 coarse-grain force field. Journal of Chemical Theory and Computation. 2021; 18(1):357–373.

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

545 **Bremer A**, Farag M, Borcherds WM, Peran I, Martin EW, Pappu RV, Mittag T. Deciphering how naturally occurring
546 sequence features impact the phase behaviours of disordered prion-like domains. Nature Chemistry. 2022;
547 14(2):196–207. https://doi.org/10.1038/s41557-021-00840-w, doi: 10.1038/s41557-021-00840-w.

548 **Brewer SH**, Vu DM, Tang Y, Li Y, Franzen S, Raleigh DP, Dyer RB. Effect of modulating unfolded state structure
549 on the folding kinetics of the villin headpiece subdomain. Proceedings of the National Academy of Sciences.
550 2005; 102(46):16662–16667. https://www.pnas.org/content/102/46/16662, doi: 10.1073/pnas.0505432102.

551 **Buhr J**, Franz F, Gräter F. Intrinsically disordered region of talin's FERM domain functions as an initial PIP2
552 recognition site. Biophysical Journal. 2023; .

553 **Bussi G**, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. Journal of Chemical Physics.
554 2007; 126(1):1–7. doi: 10.1063/1.2408420.

555 **Clavel D**, Gotthard G, von Stetten D, De Sanctis D, Pasquier H, Lambert GG, Shaner NC, Royant A. Structural anal-
556 ysis of the bright monomeric yellow-green fluorescent protein mNeonGreen obtained by directed evolution.
557 Acta Crystallographica Section D. 2016 dec; 72(12):1298–1307. https://doi.org/10.1107/S2059798316018623,
558 doi: 10.1107/S2059798316018623.

559 **Claveras Cabezudo A**, Athanasiou C, Tsengenes A, Wade RC. Scaling Protein–Water Interactions in the
560 Martini 3 Coarse-Grained Force Field to Simulate Transmembrane Helix Dimers in Different Lipid Environ-
561 ments. Journal of Chemical Theory and Computation. 2023 feb; https://doi.org/10.1021/acs.jctc.2c00950, doi:
562 10.1021/acs.jctc.2c00950.

563 **Collins PM**, Hidari KIPJ, Blanchard H. Slow diffusion of lactose out of galectin-3 crystals monitored by X-ray
564 crystallography: possible implications for ligand-exchange protocols. Acta Crystallographica Section D. 2007
565 mar; 63(3):415–419. https://doi.org/10.1107/S090744490605270X, doi: 10.1107/S090744490605270X.

566 **Cordeiro TN**, Sibille N, Germain P, Barthe P, Boulahtouf A, Allemand F, Bailly R, Vivat V, Ebel C, Barducci
567 A, Bourguet W, le Maire A, Bernadó P. Interplay of Protein Disorder in Retinoic Acid Receptor Het-
568 erodimer and Its Corepressor Regulates Gene Expression. Structure. 2019; 27(8):1270–1285.e6. doi:
569 10.1016/j.str.2019.05.001.

570 **Cornish J**, Chamberlain SG, Owen D, Mott HR. Intrinsically disordered proteins and membranes: a marriage of
571 convenience for cell signalling? Biochemical Society Transactions. 2020; 48(6):2669–2689.

572 **Das T**, Eliezer D. Membrane interactions of intrinsically disordered proteins: The example of alpha-synuclein.
573 Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics. 2019; 1867(10):879–889.

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

577 **Fagerberg E**, Månsson LK, Lenton S, Skepö M. The Effects of Chain Length on the Structural Properties of Intrin-
578 sically Disordered Proteins in Concentrated Solutions. Journal of Physical Chemistry B. 2020; 124(52):11843–
579 11853. doi: 10.1021/acs.jpcb.0c09635.

580 **Fakhree MA**, Blum C, Claessens MM. Shaping membranes with disordered proteins. Archives of biochemistry
581 and biophysics. 2019; 677:108163.

582 **Flyvbjerg H**, Petersen HG. Error estimates on averages of correlated data. The Journal of Chemical Physics.
583 1989; 91(1):461–466. doi: 10.1063/1.457480.

584 **Gillespie JR**, Shortle D. Characterization of long-range structure in the denatured state of staphylococcal nu-
585 clease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures.
586 Journal of molecular biology. 1997; 268(1):170–184.

587 **Go N**. Theoretical studies of protein folding. Annual review of biophysics and bioengineering. 1983; 12(1):183–
588 210.

589 **Gomes GNW**, Krzeminski M, Namini A, Martin EW, Mittag T, Head-Gordon T, Forman-Kay JD, Gradinaru
590 CC. Conformational Ensembles of an Intrinsically Disordered Protein Consistent with NMR, SAXS, and
591 Single-Molecule FRET. Journal of the American Chemical Society. 2020; 142(37):15697–15710. doi:
592 10.1021/jacs.0c02088.

Goretzki B, Wiedemann C, McCray BA, Schaefer SL, Jansen J, Tebbe F, Mitrovic SA, Noeth J, Donohue J, Jeffries CM, et al. Crosstalk between regulatory elements in the disordered TRPV4 N-terminus modulates lipid-dependent channel activity. bioRxiv. 2022; p. 2022–12.

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 dec; 33(6):1415–1421. https://doi.org/10.1107/S0021889800012930, doi: 10.1107/S0021889800012930.

Herzog FA, Braun L, Schoen I, Vogel V. Improved side chain dynamics in MARTINI simulations of protein–lipid interfaces. Journal of chemical theory and computation. 2016; 12(5):2446–2458.

Howard SB, Twigg PJ, Baird JK, Meehan EJ. The solubility of hen egg-white lysozyme. Journal of Crystal Growth. 1988; 90(1-3):94–104.

Idowu SM, Gautel M, Perkins SJ, Pfuhl M. Structure, Stability and Dynamics of the Central Domain of Cardiac Myosin Binding Protein C (MyBP-C): Implications for Multidomain Assembly and Causes for Cardiomyopathy. Journal of Molecular Biology. 2003; 329(4):745–761. https://www.sciencedirect.com/science/article/pii/S002228360300425X, doi: https://doi.org/10.1016/S0022-2836(03)00425-X.

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 apr; 126(18):5879–5896.

Javanainen M, Martinez-Seara H, Vattulainen I. Excessive aggregation of membrane proteins in the Martini model. PLOS ONE. 2017 nov; 12(11):e0187936. https://doi.org/10.1371/journal.pone.0187936.

Javanainen M, Martinez-Seara H, Vattulainen I. Excessive aggregation of membrane proteins in the Martini model. PloS one. 2017; 12(11):e0187936.

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.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596(7873):583–589. https://doi.org/10.1038/s41586-021-03819-2, doi: 10.1038/s41586-021-03819-2.

Jussupow A, Kaila VRI. Effective Molecular Dynamics from Neural Network-Based Structure Prediction Models. Journal of Chemical Theory and Computation. 2023 mar; https://doi.org/10.1021/acs.jctc.2c01027, doi: 10.1021/acs.jctc.2c01027.

Jussupow A, Messias AC, Stehle R, Geerlof A, Solbak SMØ, Paissoni C, Bach A, Sattler M, Camilloni C. The dynamics of linear polyubiquitin. Science Advances. 2023 mar; 6(42):eabc3786. https://doi.org/10.1126/sciadv.abc3786, doi: 10.1126/sciadv.abc3786.

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

Kjaergaard M, Kragelund BB. Functions of intrinsic disorder in transmembrane proteins. Cellular and Molecular Life Sciences. 2017; 74:3205–3224.

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.

Komander D, Reyes-Turcu F, Licchesi JDF, Odenwaelder P, Wilkinson KD, Barford D. Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. EMBO reports. 2009 may; 10(5):466–473. https://doi.org/10.1038/embor.2009.55, doi: https://doi.org/10.1038/embor.2009.55.

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 may; 17(5):3088–3102. https://doi.org/10.1021/acs.jctc.0c00507, doi: 10.1021/acs.jctc.0c00507.

Larsen AH, Pedersen MC. Experimental noise in small-angle scattering can be assessed using the Bayesian indirect Fourier transformation. Journal of Applied Crystallography. 2021 oct; 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. 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.

Lin YH, Qiu DC, Chang WH, Yeh YQ, Jeng US, Liu FT, Huang Jr. The intrinsically disordered N-terminal domain of galectin-3 dynamically mediates multisite self-association of the protein through fuzzy interactions. Journal of Biological Chemistry. 2017; 292(43):17845–17856. https://www.sciencedirect.com/science/article/pii/S0021925820330441, doi: https://doi.org/10.1074/jbc.M117.802793.

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 apr; 17(4):2513–2521. https://doi.org/10.1021/acs.jctc.0c01253, doi: 10.1021/acs.jctc.0c01253.

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.

Marrink SJ, Tieleman DP. Perspective on the Martini model. Chemical Society Reviews. 2013; 42(16):6801–6822.

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.

McKnight CJ, Matsudaira PT, Kim PS. NMR structure of the 35-residue villin headpiece subdomain. Nature Structural Biology. 1997; 4(3):180–184. https://doi.org/10.1038/nsb0397-180, doi: 10.1038/nsb0397-180.

Michie K, Kwan A, Tung CS, Guss J, Trewhella J. A Highly Conserved Yet Flexible Linker Is Part of a Polymorphic Protein-Binding Domain in Myosin-Binding Protein C. Structure. 2016; 24(11):2000–2007. https://www.sciencedirect.com/science/article/pii/S0969212616302684, doi: https://doi.org/10.1016/j.str.2016.08.018.

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nature methods. 2022; 19(6):679–682.

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 low-complexity domain disrupts phase separation, aggregation, and toxicity . The EMBO Journal. 2017; 36(20):2951–2967. doi: 10.15252/embj.201696394.

691 **Monticelli L**, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI coarse-grained
692 force field: Extension to proteins. Journal of Chemical Theory and Computation. 2008; 4(5):819–834. doi:
693 10.1021/ct700324x.

694 **Moses D**, Guadalupe K, Yu F, Flores E, Perez A, McAnelly R, Shamoon NM, Cuevas-Zepeda E, Merg AD, Martin
695 EW, Holehouse AS, Sukenik S. Structural biases in disordered proteins are prevalent in the cell. bioRxiv. 2022;
696 https://www.biorxiv.org/content/early/2022/05/12/2021.11.24.469609, doi: 10.1101/2021.11.24.469609.

697 **Mylonas E**, Hascher A, Bernadó P, Blackledge M, Mandelkow E, Svergun DI. Domain conformation of tau
698 protein studied by solution small-angle X-ray scattering. Biochemistry. 2008; 47(39):10345–10353. doi:
699 10.1021/bi800900d.

700 **Nadvi N**, Michie K, Kwan A, Guss J, Trewhella J. Clinically Linked Mutations in the Central Domains of
701 Cardiac Myosin-Binding Protein C with Distinct Phenotypes Show Differential Structural Effects. Struc-
702 ture. 2016; 24(1):105–115. https://www.sciencedirect.com/science/article/pii/S0969212615004621, doi:
703 https://doi.org/10.1016/j.str.2015.11.001.

704 **Naughton FB**, Kalli AC, Sansom MS. Association of peripheral membrane proteins with membranes: Free
705 energy of binding of GRP1 PH domain with phosphatidylinositol phosphate-containing model bilayers. The
706 journal of physical chemistry letters. 2016; 7(7):1219–1224.

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

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

712 **Pesce F**, Lindorff-Larsen K. Refining conformational ensembles of flexible proteins against small-angle x-ray
713 scattering data. Biophysical journal. 2021; 120(22):5124–5135.

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

717 **Pettersen EF**, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visual-
718 ization system for exploratory research and analysis. Journal of computational chemistry. 2004; 25(13):1605–
719 1612.

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

722 **Poma AB**, Cieplak M, Theodorakis PE. Combining the MARTINI and Structure-Based Coarse-Grained Ap-
723 proaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins. Journal of Chem-
724 ical Theory and Computation. 2017 mar; 13(3):1366–1374. https://doi.org/10.1021/acs.jctc.6b00986, doi:
725 10.1021/acs.jctc.6b00986.

726 **Qi Y**, Ingólfsson HI, Cheng X, Lee J, Marrink SJ, Im W. CHARMM-GUI martini maker for coarse-grained simulations
727 with the martini force field. Journal of chemical theory and computation. 2015; 11(9):4486–4494.

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

731 **Rose PW**, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, Green RK, Goodsell DS, Prlić A, Quesada M,
732 et al. The RCSB Protein Data Bank: new resources for research and education. Nucleic acids research. 2012;
733 41(D1):D475–D482.

734 **Ryan VH**, Dignon GL, Zerze GH, Chabata CV, Silva R, Conicella AE, Amaya J, Burke KA, Mittal J, Fawzi NL. Mechanis-
735 tic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Muta-
736 tion and Arginine Methylation. Molecular cell. 2018 feb; 69(3):465–479.e7. doi: 10.1016/j.molcel.2017.12.022.

737 **Šali A**, Blundell TL. Comparative Protein Modelling by Satisfaction of Spatial Restraints. Journal of Molecu-
738 lar Biology. 1993; 234(3):779–815. http://www.sciencedirect.com/science/article/pii/S0022283683716268, doi:
739 https://doi.org/10.1006/jmbi.1993.1626.

740 **Snead D**, Wragg RT, Dittman JS, Eliezer D. Membrane curvature sensing by the C-terminal domain of complexin.
741 Nature communications. 2014; 5(1):4955.

742 **Sonntag M**, Jagtap PKA, Simon B, Appavou MS, Geerlof A, Stehle R, Gabel F, Hennig J, Sattler M. Seg-
743 mental, Domain-Selective Perdeuteration and Small-Angle Neutron Scattering for Structural Analysis of
744 Multi-Domain Proteins. Angewandte Chemie - International Edition. 2017; 56(32):9322–9325. doi:
745 10.1002/anie.201702904.

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

750 **Srinivasan S**, Zoni V, Vanni S. Estimating the accuracy of the MARTINI model towards the investigation of
751 peripheral protein–membrane interactions. Faraday Discuss. 2021; 232(0):131–148. http://dx.doi.org/10.
752 1039/D0FD00058B, doi: 10.1039/D0FD00058B.

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

756 **von Stetten D**, Noirclerc-Savoye M, Goedhart J, Gadella Jr TWJ, Royant A. Structure of a fluorescent protein from
757 *Aequorea victoria* bearing the obligate-monomer mutation A206K. Acta Crystallographica Section F. 2012 aug;
758 68(8):878–882. https://doi.org/10.1107/S1744309112028667, doi: 10.1107/S1744309112028667.

759 **Tesei G**, Martins JM, Kunze MBA, Wang Y, Crehuet R, Lindorff-Larsen K. DEER-PREdict: Software for efficient
760 calculation of spin-labeling EPR and NMR data from conformational ensembles. PLOS Computational Biology.
761 2021 jan; 17(1):e1008551. https://doi.org/10.1371/journal.pcbi.1008551, doi: 10.1371/journal.pcbi.1008551.

762 **Tesei G**, Schulze TK, Crehuet R, Lindorff-Larsen K. Accurate model of liquid–liquid phase behavior of intrinsically
763 disordered proteins from optimization of single-chain properties. Proceedings of the National Academy of
764 Sciences. 2021; 118(44). https://www.pnas.org/content/118/44/e2111696118, doi: 10.1073/pnas.2111696118.

765 **Thomasen FE**, Lindorff-Larsen K. Conformational ensembles of intrinsically disordered proteins and flexible
766 multidomain proteins. Biochemical Society Transactions. 2022 feb; p. BST20210499. https://doi.org/10.1042/
767 BST20210499, doi: 10.1042/BST20210499.

768 **Thomasen FE**, Pesce F, Roesgaard MA, Tesei G, Lindorff-Larsen K. Improving Martini 3 for Disordered and
769 Multidomain Proteins. Journal of Chemical Theory and Computation. 2022 apr; 18(4):2033–2041. https:
770 //doi.org/10.1021/acs.jctc.1c01042, doi: 10.1021/acs.jctc.1c01042.

771 **Tunyasuvunakool K**, Adler J, Wu Z, Green T, Zielinski M, Žídek A, Bridgland A, Cowie A, Meyer C, Laydon A, Ve-
772 lankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko
773 A, et al. Highly accurate protein structure prediction for the human proteome. Nature. 2021; 596(7873):590–
774 596. https://doi.org/10.1038/s41586-021-03828-1, doi: 10.1038/s41586-021-03828-1.

775 **Varadi M**, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, et al.
776 AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence
777 space with high-accuracy models. Nucleic acids research. 2022; 50(D1):D439–D444.

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

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

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

787 **Webb B**, Sali A. Comparative protein structure modeling using MODELLER. Current protocols in bioinformatics.
788 2016; 54(1):5–6.

789 **Yamamoto E**, Kalli AC, Akimoto T, Yasuoka K, Sansom MS. Anomalous dynamics of a lipid recognition protein
790 on a membrane surface. Scientific reports. 2015; 5(1):18245.

Zeno WF, Baul U, Snead WT, DeGroot AC, Wang L, Lafer EM, Thirumalai D, Stachowiak JC. Synergy between intrinsically disordered domains and structured proteins amplifies membrane curvature sensing. Nature communications. 2018; 9(1):4152.

Zerze GH. Optimizing the Martini 3 Force Field Reveals the Effects of the Intricate Balance between Protein–Water Interaction Strength and Salt Concentration on Biomolecular Condensate Formation. Journal of Chemical Theory and Computation. 2023 apr; https://doi.org/10.1021/acs.jctc.2c01273, doi: 10.1021/acs.jctc.2c01273.