

Martini3-IDP: improved Martini 3 force field for disordered proteins

Received: 12 November 2024

Accepted: 14 March 2025

Published online: 24 March 2025

Liguo Wang¹, Christopher Brasnett¹, Luís Borges-Araújo^{2,3},
Paulo C. T. Souza^{2,3} & Siewert J. Marrink¹✉

Coarse-grained (CG) molecular dynamics (MD) is widely used for the efficient simulation of intrinsically disordered proteins (IDPs). The Martini model, one of the most popular CG force fields in biomolecular simulation, was reported to yield too compact IDP conformations, limiting its applications. Addressing this, we optimized the bonded parameters based on fitting to reference simulations of a diverse set of IDPs at atomistic resolution, resulting in a Martini3-based disordered protein model coined Martini3-IDP. This model leads to expanded IDP conformations, greatly improving the reproduction of the experimentally measured radii of gyration. Moreover, contrary to ad-hoc fixes based on scaling of protein-protein or protein-water interactions, Martini3-IDP keeps the overall interaction balance underlying Martini 3. To validate that, we perform a comprehensive testing including full-length multidomain proteins, IDP-lipid membrane binding and IDP-small molecule binding, confirming its ability to successfully capture the complex interplay between disordered proteins and diverse biomolecular components. Finally, the recently emerging concept of biomolecular condensate, through liquid-liquid phase separation, was also reproduced by Martini3-IDP for a number of both homotypic and heterotypic systems. With the improved Martini3-IDP model, we expand the ability to simulate processes involving IDPs in complex environments, at spatio-temporal scales inaccessible with all-atom models.

In eukaryotic proteomes, more than 30% of the residues are classified as intrinsically disordered^{1,2}. And in human proteomes, around 70% of proteins are reported to possess disordered regions of 30 residues length or even longer³. In contrast to conventional well-structured folded proteins, intrinsically disordered protein (IDPs) or disordered regions (IDRs) exist in an ensemble of interconverting distinct conformations without an overall stable 3D structure. Due to this conformational plasticity, IDPs and IDRs are capable of mediating multivalent, malleable molecular interactions with diverse partners and function in a wide range of cellular processes, including gene transcription, translation, cell signaling, nuclear transport and local organization^{4–10}.

Molecular dynamics (MD) is an effective tool for characterizing the conformational ensembles and resulting microscopic behavior of IDPs and IDRs. Limited by current computing speed, high-accuracy atomistic resolution simulations of IDP/IDRs remain challenging, because coverage of larger spatio-temporal scales compared to conventional globular proteins is required to adequately sample the configurational space. Consequently, coarse-grained (CG) models, which represent groups of atoms as single beads, are widely used as a computationally efficient alternative to atomistic models while still capturing essential structural and dynamic features¹¹.

One of the most popular CG models is Martini^{12,13}, a generic force field for simulation of (bio)molecular interactions. However, recent

¹Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9747AG Groningen, The Netherlands. ²Laboratoire de Biologie et Modélisation de la Cellule, CNRS, UMR 5239, Inserm, U1293, Université Claude Bernard Lyon 1, Ecole Normale Supérieure de Lyon, Lyon, France. ³Centre Blaise Pascal de Simulation et de Modélisation Numérique, Ecole Normale Supérieure de Lyon, Lyon, France. ✉e-mail: s.j.marrink@rug.nl

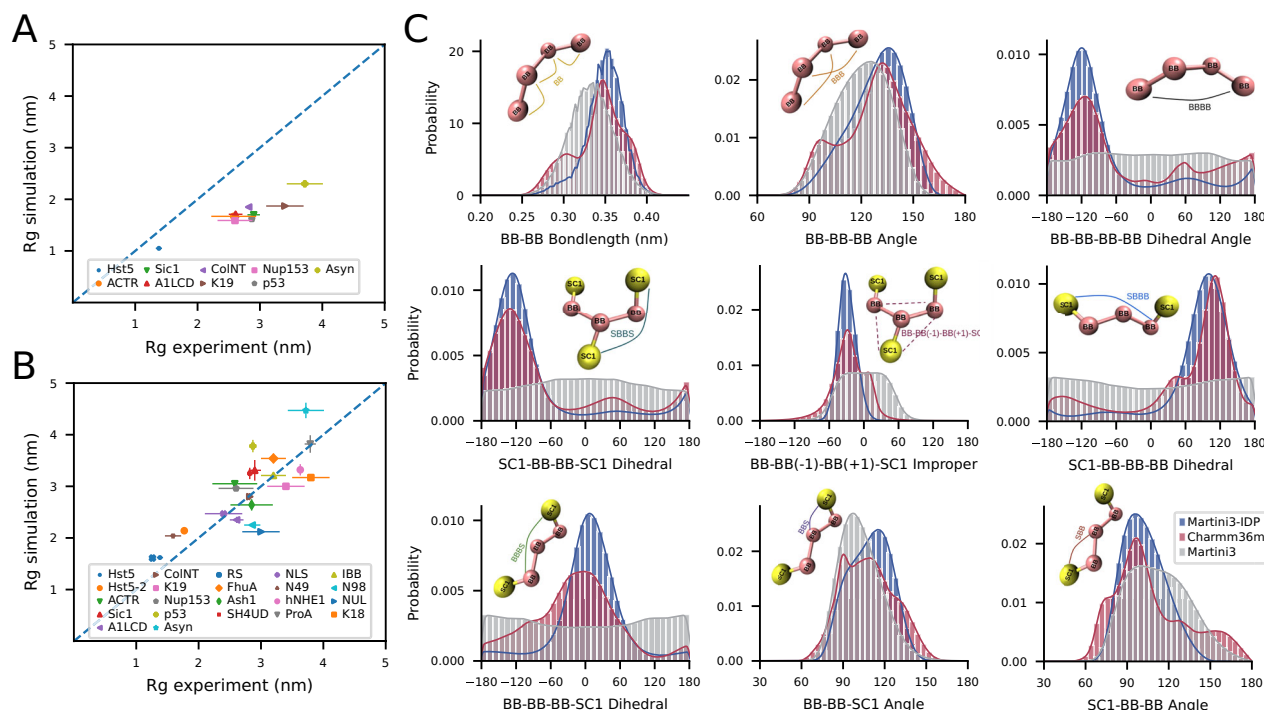


Fig. 1 | IDP optimization in Martini 3. Mean Rg calculated from simulations plotted against experimental Rg for **A** standard Martini 3 (MAE = 1.058 nm, PCC = 0.948) and **B** Martini3-IDP (MAE = 0.387 nm, PCC = 0.803). Rg data are presented as mean values \pm statistical error estimates, error bars of experimental Rg values were determined by a fitting error from the Guinier fit and error bars of simulations represent the standard deviation of 5 blocks from blocking analysis. **C** Comparison of bonded distributions involving protein backbone (BB) and first side chain (SC1) beads between Martini3, Martini3-IDP and CHARMM36m. KL divergence between

bonded distributions are provided in Supplementary Table 10. The distributions of CHARMM36m were obtained by mapping the AA trajectory to Martini resolution. Due to the different nature of Gly residue, BB-BB-BB-SC1 and SC1-BB-BB-BB dihedral angles involving Gly were excluded and are provided separately in Supplementary Fig. 4. BB represents the BB-BB bond length; BBB, BBS and SBB represent the BB-BB-BB, BB-BB-SC1 and SC1-BB-BB angles; BBBB, SBSB, SBBB and BBBS represent the BB-BB-BB-BB, SC1-BB-BB-SC1, SC1-BB-BB-BB and BB-BB-BB-SC1 dihedral angle, respectively.

work indicated that the latest version, Martini 3, results in overly compact conformations across a wide set of IDPs and multi-domain proteins with IDR linkers^{14–16}. To recover accurate radii of gyration compared to experimental data, different strategies have been proposed, including shifting backbone-water interactions via virtual sites¹⁷ or complete rescaling of protein-water or protein-protein interactions^{14,18}, analogous to approaches used in the atomistic simulation field to expand IDP conformations¹⁹. Although such pragmatic fixes can be used in certain applications, they offset the overall balance of interactions with other biomolecules in the Martini model. Moreover, it has been shown that folded proteins in Martini 3 are less sticky after alleviating the protein aggregation problem in Martini 2 and protein dimerization is relatively weak in aqueous solution^{20–22}. Rescaling of protein-protein or protein-water interactions for folded proteins would therefore be required in directions opposite to IDPs, making this strategy undesirable.

In this work we aimed to identify the primary cause of the conformational collapse of IDPs in Martini 3. To this end, in addition to a well-established benchmark previously published¹⁷, we ran microsecond length simulations on a new set of IDPs, using Martini 3 and an all-atom (AA) model for comparison. By analyzing this combined dataset, we found a substantial discrepancy in the locally accessible configurations sampled by Martini 3 compared to the reference atomistic force field. By optimizing the protein bonded parameters, targeting the configurational distributions obtained with atomistic resolution^{23,24} as a reference, we developed a Martini 3 disordered protein model, coined Martini3-IDP. This model samples more expanded IDP conformations, aligning better with atomistic simulations and performing well in reproducing experimental radii of gyration. Moreover, Martini3-IDP is well integrated with the current Martini

framework and is available for general applications, offering an advantage over other specifically trained CG force fields for IDPs^{25–27}. To validate this, we measured overall conformational metrics, including radius of gyration, across a diverse set of multidomain proteins finding good agreement with experimental values. Additionally, the interactions between IDPs and several other molecule types in the extensive chemical space covered by the Martini framework—such as lipid membranes and small-molecule ligands—were validated. Finally, for both homotypic and heterotypic IDP-based systems, biomolecular condensate formation via liquid-liquid phase separation (LLPS) was also reproduced by Martini3-IDP with near quantitative accuracy.

Results

Martini 3 samples collapsed conformations of IDPs

To analyze the conformational behavior of IDPs in Martini 3, we chose 9 typical IDPs (Supplementary Table 1), covering diverse systems from the recent benchmark work which suggested too compact IDP conformations in Martini 3¹⁴. Each IDP was simulated in excess solvent for 10–25 μ s, which is sufficient for convergence of the radius of gyration (Rg) (Supplementary Fig. 14). As shown in Fig. 1A and consistent with previous reports, Martini 3 recovers overly compact IDP conformations across the whole dataset, as seen by the underestimated Rg compared to values derived from small-angle X-ray scattering (SAXS) data, with a mean absolute error (MAE) of 1.058 nm. Interestingly, the collapsed IDPs in Martini 3 still show a good overall correlation with experimental data (Pearson correlation coefficient PCC = 0.948), suggesting that the shortcoming is not sequence-dependent but rather originates from a more generic feature of the Martini 3 IDP model. Together with the notion that Martini 3 is considered less sticky, with key protein-protein interactions being on the weak side rather than too

strong^{20–22}, increasing the effective hydration via globally increasing protein-water or decreasing protein-protein interactions may not provide a generic and transferable solution for all cases. Modifying bonded interactions, one popular approach in atomistic IDP model development, presents an unexplored alternative in Martini CG models that could address these challenges. This prompted us to look into the bonded interactions, which affect the local accessible configurations and in principle can also have a large impact on the global dimensions of IDPs. A similar strategy was recently applied to improve the conformational ensembles of IDPs modeled with the one-bead-per-residue Mpipi force field²⁵.

To assess to what extent local configurations are realistically sampled, we performed additional reference AA simulations based on the CHARMM36m force field for a set of 4 IDPs (RS-peptide, FG-peptide, HEWL19, and polyQ30, details in Supplementary Table 6). The resulting 5 μ s trajectories were augmented by a set of overall 1.4 μ s atomistic trajectories for 12 additional IDPs from a recent work¹⁷. Together, these 16 IDP atomistic trajectories were mapped to Martini resolution to examine the effective angle and dihedral distributions. Comparing the results revealed a considerable discrepancy between Martini 3 and CHARMM36m distributions, especially those concerning the backbone dihedral and side chain orientation (Fig. 1C). This finding is not surprising, as in Martini, no dihedral potential is defined for coil secondary structure regions due to its broad distribution. For other secondary structure elements, backbone dihedrals are included in the model, while no dihedral between backbone-side chains are defined for any of the secondary structure elements. Side chain orientation in the Martini 3 protein models can be improved by the use of -scFix flag in Martinize2²⁸, but these are dependent on an atomistic reference structure and are not suitable for IDPs. Presumably, the lack of dihedrals and the consequential deviations in bonded parameter distributions could undermine the global IDP conformational sampling, as the model could reach conformations geometrically impossible. For example, one may expect that unrestricted sampling of side chain orientations would facilitate neighboring side chains approaching each other to obtain a more favorable interaction energy, causing protein chain collapse as a consequence.

Bonded parameter fitting expands IDP conformations

Due to the broad distribution of bonded parameters in coil secondary structures or IDPs, previous Martini versions did not define specific angle and dihedral potentials for these regions. This simplification was made for convenience, as similar approaches have been applied to other flexible components in Martini, such as lipid tails, which also lack dihedral definitions. While this approach simplifies the model and improves numerical stability, it results in a less accurate representation of the energy landscape. As shown in Fig. 1C, distributions extracted from atomistic trajectories of a wide variety of IDPs reveal a more detailed and complex energy landscape. This suggests that developing improved bonded parameters for disordered proteins, based on fitting to these atomistic configurations, could enhance the accuracy of the model.

To start, the BB-BB-BB-BB dihedral along the backbone beads (BBs) shows a dominant distribution peak around -120° and a local peak around 60° reflecting the presence of transient helical motifs in the IDP conformational ensemble. This contrasts heavily with the relatively smooth and uniform distribution obtained with Martini 3. To obtain more insight into the residue specificity of these distributions, we considered dihedrals with Gly and Pro separately, being two unique amino acids in protein secondary structure formation and treated differently from the other 18 amino acids in the CHARMM36m grid-based energy correction map parameterization²⁹. The four-residue length subsequences with different Gly or Pro content were extracted from the whole database (Supplementary Fig. 1), revealing that the subsequences with different Pro content showed similar overall BB-BB-

BB-BB distribution characteristics as the full set, however, the distribution peak of subsequences with different Gly content shifted significantly with Gly content increasing. Clearly, subsequences with 3 Gly showed a distinctly different distribution compared with the overall BB-BB-BB-BB backbone dihedral (GGGG subsequence was rarely sampled in the database and so not considered). To better understand the behavior of these high-Gly-content subsequences, several GGGXGGG peptides (X: A, Q, P, Y, R, D) as well as a Gly10 toy peptide were designed. AA simulations of these model peptides revealed that the nature of the inserted residue did not affect the BB-BB-BB-BB dihedral distribution much. In contrast, the residue insertion position played an important role (Supplementary Fig. 2). Based on this finding, we classified Gly-rich sequences into three position patterns: GGGX/XGGG, GGXG and GXGG. Besides, the GGGG subsequence extracted from the Gly10 peptide was found to have a similar distribution as GGGX/XGGG (Supplementary Fig. 3A), thus was merged into this pattern class.

In summary, the atomistic trajectory mapping suggests that disordered proteins exhibit a relatively general BB-BB-BB-BB backbone dihedral distribution with a well-defined maximum at approximately -120° . These are not at all captured by Martini3. In addition, we identified some high-Gly-content subsequence patterns (making up $\sim 3\%$ protein sequence in our AA resolution data) featuring distinct BB-BB-BB-BB backbone dihedral distributions, which is in line with the unique role of the GGG motif in IDP backbone dynamics reported recently³⁰. In principle, other residue patterns with varying secondary structure formation propensity could exist, and explicit description of them could further improve backbone conformational sampling, including local transient helicity prevalent in IDPs. However, it is impractical to consider all 20^4 combinations, even if restricted to known IDP-like sequences. As such, only the three specific high-Gly-content subsequence patterns were explicitly taken into account to refine our model.

To improve the dihedral conformations accessible with Martini, we incorporate both a generic BB-BB-BB-BB dihedral potential and a Glycine-specific dihedral potential in Martini3-IDP, using the Gromacs type 9 dihedral function that merges multiple proper dihedral functions to match complicated distributions. To compare to the AA results, we simulated a larger benchmark set incorporating 22 IDPs (Supplementary Table 2). With these additional potentials, Martini3-IDP markedly improves the backbone torsional behavior, as can be seen from the dihedral distributions being more accurately reproduced when compared to the reference simulations (Fig. 1C).

In parallel, the backbone BB-BB bond length and BB-BB-BB angle were slightly shifted from 0.35 nm and 127° to 0.36 nm and 137° , respectively, to better capture these distributions (Fig. 1C). Bond lengths involving Pro residues were also improved, and made dependent on the flanking residue to account for the mapping effect: 0.321, 0.305, and 0.360 nm for PP, PX, and XP backbone BB-BB bonds, respectively (Supplementary Fig. 3B).

Next, we examined side chain orientation. As side-chain (SC) orientation was previously noted as important²⁸, we observed that neighboring side chains in disordered proteins tend to adopt relatively stable orientations, with the SC1-BB-BB-SC1 dihedral distribution (SC1 being the side chain bead directly attached to the BB in the Martini model) showing a major peak around -120° corresponding to a preference for a near antiparallel orientation (Fig. 1C). While this SC1-BB-BB-SC1 dihedral distribution could be captured by adding a dihedral potential (similar to what is done for folded Martini proteins using the -scFix option²⁸), the coupling of this dihedral with other backbone dihedrals requires a comprehensive approach that considers all relevant dihedrals together.

The distribution of the BB-BB(-1)-BB(+1)-SC1 dihedral, (BB(+1) and BB(-1) beads denoting the following and preceding backbone beads, respectively), is also too broad in Martini 3 compared to the sharp

orientation observed in reference AA trajectories. Although BB beads and C α atoms are not exactly the same in CG and AA models, their positions are close enough for this dihedral to obey a similar tetrahedral configuration of *sp*³-hybrid carbon atoms observed in the IDP backbones (Fig. 1C, Supplementary Fig. 5C). Thus a residue-dependent BB-BB(-1)-BB(+1)-SC1 improper dihedral (Gromacs type 2 harmonic dihedral function) was introduced in Martini to capture this side chain orientation feature (Supplementary Fig. 5A). This intrinsic BB-BB(-1)-BB(+1)-SC1 improper dihedral also restrained neighboring SC1-BB-BB-SC1 dihedral configurations through coupling. Taking the IDP Sic1 as an example, the SC1-BB-BB-SC1 distribution characteristics were already captured without the need for an explicit definition of the SC1-BB-BB-SC1 dihedral (Supplementary Fig. 5B). Thus, in the current work, only a weak SC1-BB-BB-SC1 potential was added on the basis of the BB-BB(-1)-BB(+1)-SC1 improper dihedral definition to further enhance the antiparallel orientation of neighboring side chains (Fig. 1C). In the same way, the BB-BB-BB-SC1 and SC1-BB-BB-BB dihedrals were improved due to the existing coupling with other dihedrals. As shown in Fig. 1C, SC1-BB-BB(Gly)-BB, SC1-BB-BB(noGly)-BB and BB-BB(noGly)-BB-SC1 were all mostly well-captured without explicit definition, except for a weak BB-BB(Gly)-BB-SC1 dihedral potential added to slightly optimize the fitting. Finally, BB-BB-SC1 and SC1-BB-BB angle potentials were optimized in a residue-dependent manner (Fig. 1C, Supplementary Fig. 6).

Overall, the improved bonded parameters on the basis of standard Martini 3 highly improved IDP conformational sampling in Martini, as evidenced by the improved match of the distributions between Martini and CHARMM36m (Fig. 1C and Supplementary Table 10). Importantly, the new bonded parameters also led to a significant expansion of the IDP conformations. In a large dataset containing 22 widely-used IDPs (Supplementary Table 2), one of the most diverse IDP datasets in recent IDP model development as we know, it could predict the Rg within a MAE of 0.387 nm and a Pearson's correlation coefficient of 0.803, comparable to recently developed IDP-specific one-residue-per-bead CG models (Fig. 1B)^{25–27}.

Moreover, since the experimental Rg is an ensemble-averaged value with a grade of ambiguity, we aimed for a more direct comparison of the conformational ensembles between our proposed Martini3-IDP and higher resolution models, which are available for a few isolated IDPs simulated for tens of μ s with the Amber99SB-disp/CHARMM36m/a03ws atomistic force field³¹ (see Methods). Principal component analysis (PCA) was used to project the conformational ensembles from different force fields into the same 2D coordinates (Supplementary Fig. 7). It is noted that some atomistic ensembles are slightly compact when compared with experimental Rg data (for instance, Rg of ACTR in the Amber99SB-disp ensemble is 2.13 nm, smaller than 2.58 nm derived from SAXS data), while the ensembles from our Martini3-IDP simulation are more expanded in these three systems, which could amplify the conformational discrepancy between them. In spite of this discrepancy, our analysis confirmed that standard Martini 3 only sampled a limited conformational space compared with both the atomistic ensemble and the experimentally derived NMR ensemble from the PED dataset, consistent with the IDP collapse observed in Martini 3 before. Although somewhat inconclusive due to sampling limitations at the AA level, these data provide further evidence that the bonded parameter development strategy underlying Martini3-IDP is able to generate more realistic IDP conformations while recovering IDP expansion.

Multidomain proteins with the Martini3-IDP force field

Multidomain proteins (MDPs), consisting of both IDRs and folded domains, make up a large portion (around 50%) of the proteomes in eukaryotic cells³². MDPs are particularly versatile in their response to the environment: on the one hand, enabling multivalent, malleable molecular recognition similarly to IDPs, on the other hand, allowing

specific interactions through the folded domains. Moreover, the IDRs and folded domains are not behaving isolated from each other. Folded domains could affect the dynamics of disordered tails by steric effects or surface chemical makeup^{33,34}; IDRs could also influence the function or even conformation of folded domains, either by fuzzy interactions or by exerting an entropic force^{35–39}. For example, the folded domain in the RNA-binding protein hnRNPA1 could attract its IDR tail through surface charged patches and affect the conformation of the IDR tail²⁵. Likewise, an entropic-driven force from IDR loop dynamics was shown to inhibit the dimerization of two tethered folded domains^{35,38}. Moreover, IDRs could even engage in attractive interactions with its adjacent domains as competitive inhibitors or supply additional attractive interactions with partner proteins for enhanced molecular recognition^{36,37,39}.

As a general estimation for MDPs internal interplay and global dimensions, the radius of gyration of 9 diverse MDPs were measured by Martini3-IDP (see Methods section and Supplementary Table 3). To this end, the MDPs were simulated for 20–30 μ s, using a setup with a single protein in aqueous solution, matching ionic strength and protein concentration to the experimental SAXS reference data^{15,39–43}. Simulation convergence was assessed using blocking analysis of the Rg (Supplementary Figs. 15,18). As shown in Fig. 2A, the overall Rg measurements in Martini3-IDP agree well with the experimental values, suggesting that the balance between folded-folded, folded-IDR and IDR-IDR interactions is adequately described with the improved model.

A number of more specific interplay mechanisms between IDRs and folded domains were also validated (see Methods). As indicated in Fig. 2B, we could reproduce the trends in the salt-dependent IDR tail release of the MDP hnRNPA1. These data were obtained by performing a 15 μ s simulation of the hnRNPA1 protein in aqueous solution at varying salt concentration. With salt concentration increasing, the Rg of hnRNPA1 increased (Fig. 2B), whereas the contact density between the folded domain and the tail decreased (Supplementary Fig. 8), consistent with the electrostatic attraction mode reported in the experiment²⁴.

Furthermore, we tested to what extent non-specific entropic forces of IDRs could be captured with Martini3-IDP. As mentioned above, disordered proteins in the standard Martini 3 force field suffered from conformation collapse, and their conformational entropy was limited. The expanded disordered protein conformations in Martini3-IDP should, in principle, be able to describe these entropic effects better. We focus on the consensus-designed alpha-helix tetratricopeptide repeats (CTPRs), which are able to form stable dimers with their dimerization strength tuned by insertion of disordered loops between adjacent repeats²⁶. In the current study, proteins consisting of two CTPRs separated either by a short linker (CTPR2a) or a 55 residue long loop (CTPR2-Loop10) were considered. According to chemical-induced equilibrium denaturation measurements²⁶, the addition of Loop10 decreases protein dimer stability by -6.86 kJ/mol. Based on brute force MD simulations of both these constructs, the dimerization state fraction of the two repeat units was used to calculate the dissociation constant K_D (see Eq. 1 in Methods). In line with the experimental evidence, we found that the dimerization of CTPR2 was destabilized after Loop 10 insertion (K_D increasing from 0.30 ± 0.03 mM to 2.25 ± 0.59 mM at 303 K, for example, see Supplementary Table 8). This -7.4 fold dissociation increase equals 5.0 kJ/mol dimerization stability reduction, reasonably close to the experimental stability loss of -6.86 kJ/mol. Our data suggest that, with the conformation expansion in Martini3-IDP, the conformational entropy of IDR linkers can be reproduced.

Beyond that, fuzzy bindings mediated by IDRs are ubiquitous across molecular recognitions. For example, IDRs were thought to be able to enhance partner contact by a so-called fly-casting

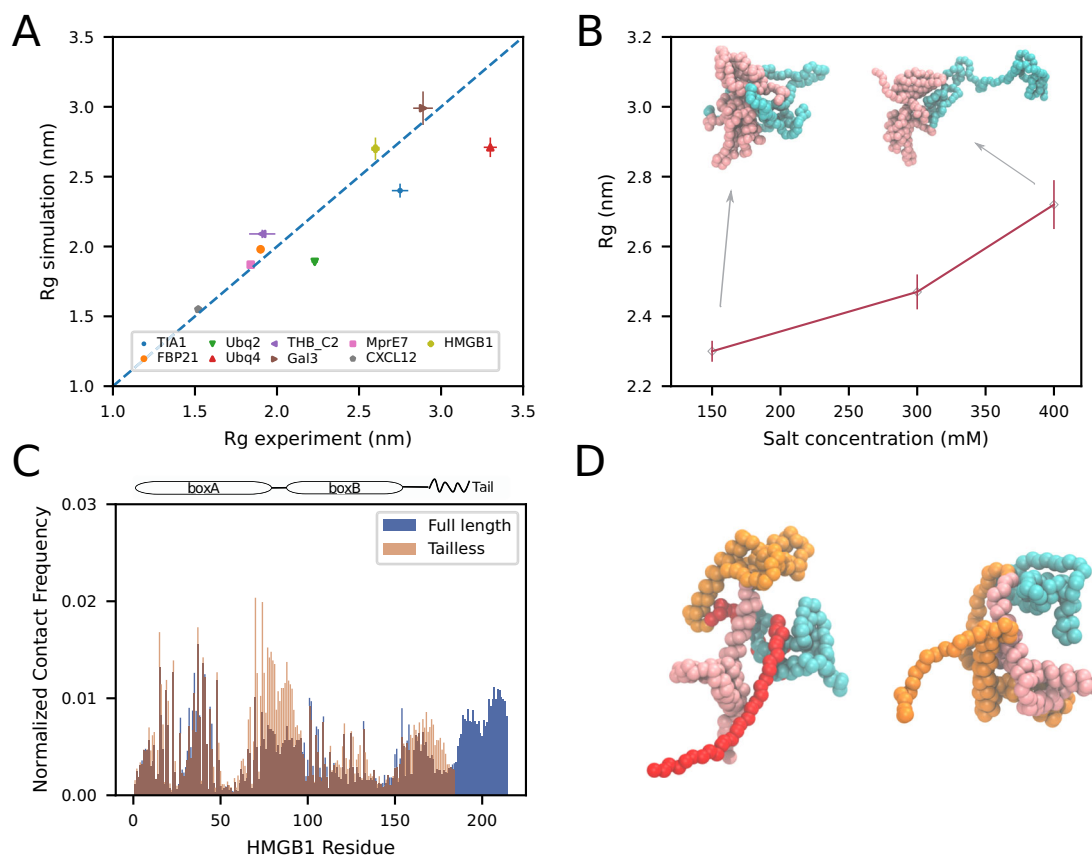


Fig. 2 | Martini3-IDP performance in MDP systems. A Mean R_g calculated from Martini3-IDP simulations plotted against experimental R_g . R_g data are presented as mean values \pm statistical error estimates, error bars of experimental R_g values were determined by a fitting error from the Guinier fit and error bars of simulations represent the standard deviation of 5 blocks from blocking analysis. **B** Full-length hnRNPA1 protein mean R_g calculated from Martini3-IDP simulations under different salt concentrations. R_g data are presented as mean values \pm statistical error estimates, error bars of simulations represent the standard deviation of 5 blocks from blocking analysis. Snapshots of the salt concentration dependent behavior of full length hnRNPA1 protein are shown. Increased electrostatic screening undermines the electrostatic attraction between RRM and LCD region and releases the LCD tail of hnRNPA1. Residues of the RRM region are colored as

pink; the LCD region is colored as cyan. **C** Normalized contact frequency along HMGB1 sequence (averaged across three simulations). The integrated contact population of boxA residue 70–90 region equals 25.8% in tailless HMGB1, significantly larger than the 12.3% in full-length HMGB1, although the contact population of the residue 150–180 region in box B in tailless HMGB1 is also slightly increased to 17.4% compared to 12.7% in presence of acid tails. This data shows that in the absence of the HMGB1 acid tail, this hetero-complex CXCL12-HMGB1 has a stronger binding preference for the HMGB1 boxA domain. **D** Snapshots of full length (left) and tailless (right) HMGB1 and CXCL12 complex. BoxA in HMGB1 is colored pink, the boxB region is colored orange, and the acid tail is red; CXCL12 is colored cyan. The acid tail interferes with the binding between boxA and CXCL12.

mechanism^{44,45}, two intrinsically disordered human proteins histone H1 and its nuclear chaperone prothymosin- α could associate in a fuzzy-complex with picomolar affinity⁴⁶. Similarly, the acid disordered tail of the HMGB1 protein (alarmin High Mobility Group Box 1) was revealed to change the protein complex conformation of HMGB1-CXCL12 (C-X-C motif ligand 12) by fuzzy interactions with CXCL12³⁰. To validate Martini 3-IDP for these kinds of interactions, we simulated the protein complex formed by CXCL12 and HMGB1 both with, and without, the disordered tail domain. Based on extensive 15 μ s triplicate simulations, we find that the acid tail enhances the association affinity (K_D 74.7 ± 20.6 μ M and 200.1 ± 83.2 μ M for full length and tailless HMGB1, shown in Supplementary Table 8), in qualitative agreement with the experimentally determined dissociation constants based on micro-scale thermophoresis measurements (K_D 31.8 μ M and 81.4 μ M for full length and tailless HMGB1, respectively³⁰). Moreover, Martini3-IDP simulations also captured the complex state change driven by the HMGB1 acid tail. Based on NMR data³⁰, in the presence of the acid tail HMGB1 would weakly associate with CXCL12 by promiscuous interactions across the whole HMGB1 protein sequence. In contrast, in the absence of the tail, the complex was found relatively stable with preferred binding on the HMGB1 boxA domain (residue 10–90). In the

Martini3-IDP simulation of this hetero-complex, binding at the boxA domain was indeed reduced in the presence of the acid tail compared to the binding mode of tailless HMGB1 (Fig. 2C), in line with the NMR data.

Taken together, the tests in this section demonstrate that Martini3-IDP could take advantage of the already validated Martini 3 folded protein model and capture the diverse, specific and unspecific, interplays between IDRs and folded domains, providing an opportunity for efficient simulation of MDPs at near-atomic resolution.

Probing IDP-ligand binding with Martini3-IDP model

Multivalent and broad binding interfaces enable IDPs to function as hubs of cellular interactomes, which makes its dysregulation associated with major diseases such as occurrence of cancers and neurodegenerative diseases^{47,48}. Although tempting, IDPs are often viewed as undruggable targets due to a lack of stable 3D conformations. Conventional structure-based drug design methods are typically unsuitable for targeting IDPs, as these rely on ligands well fitted in particular binding pockets. Recently, ensemble docking has been used to circumvent this problem^{48,49}. Instead of choosing a single conformation as target, rather an ensemble of representative and well populated

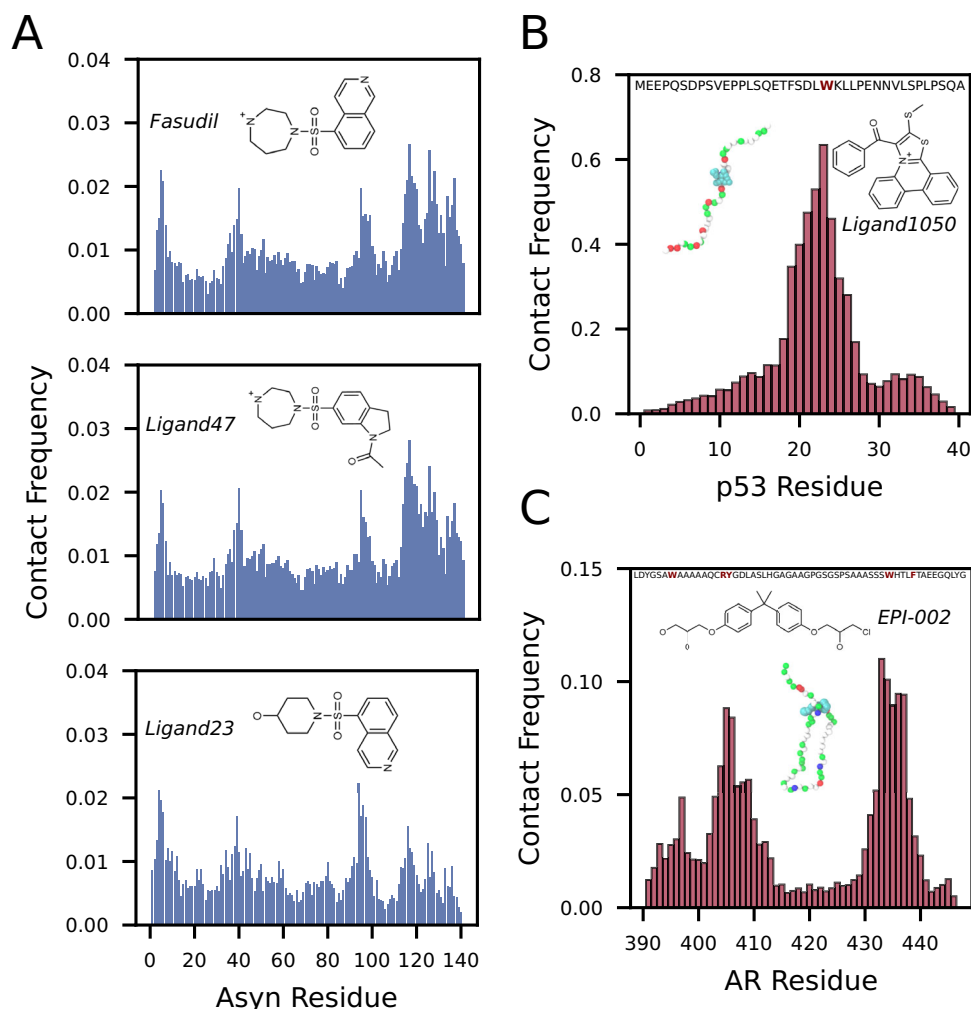


Fig. 3 | Martini3-IDP performance in IDP-ligand systems. Per-residue contact frequency between IDP and small molecule ligands in **A** Asyn **B** p53 TAD and **C** AR NTD systems. Contact frequency was calculated as the fraction of contact state during simulations. Illustrative MD snapshots of IDP-ligand binding in p53 TAD and

AR NTD systems are shown as insets. Residues are color-coded white for apolar, green for uncharged polar, red for anionic, and blue for cationic. Small molecule ligands are colored cyan. Key residues during binding are red-colored in the protein sequence.

conformations are selected for molecular docking, enabling computational virtual screening of IDPs. Creating these ensembles and optimizing hit molecules after first-round virtual screening, however, is a computationally demanding task, making AA models less well suited. The large chemical space covered and highly improved IDP conformational sampling in Martini3-IDP provide a CG level solution for small molecule design with much cheaper resource cost and potentially reasonable accuracy judged from successful Martini-based drug binding assays considering folded proteins^{48,50–53}. To test the validity of Martini3-IDP in this field, three typical IDP-ligand binding systems were chosen for validation in this study^{49,54,55}.

First, we considered ligand binding to alpha-synuclein (Asyn). For reference, making use of the special-purpose simulation machine Anton, a near-ms timescale AA simulation for Asyn-ligand binding is available. In this long-time scale AA simulation, the binding mode and affinity difference among several analog candidates was captured⁵⁴. Three extensively tested ligands (Fasudil, ligand 47 and ligand 23) were also tested with Martini3-IDP. Based on triplicate 50 μ s simulations, we observe frequent binding/unbinding events ('dynamic shuttling'), with overall binding affinities in the mM range in near quantitative agreement with the results from the AA simulations (Supplementary Table 9). In particular, we predict the same order of binding strength, with ligand 47 being the strongest binder, and ligand 23 the weakest.

Moreover, as shown in Fig. 3A, Fasudil and ligand 47 have a binding preference to the C-terminal region, while the binding preference of analog ligand 23 shifts to the N-terminal region, in line with the AA simulations as well as with NMR data⁵⁴.

As a second showcase, we consider the interaction between the transactivation domain (TAD) of p53 and its inhibitor ligand 1050. The p53 TAD binds several negative regulators that suppress its ability to activate transcription. Ligand 1050 from the SPECS library was discovered by hierarchical ensemble docking and phenotype screening to restore wild-type p53 function in cancer cells⁴⁹. NMR chemical shift perturbations data and AA simulation showed that ligand 1050 predominantly binds to a region around a specific tryptophan (W23)⁴⁹. Based on 30 μ s simulations, we could replicate this behavior, as shown in Fig. 3B. The ligand binding also reduced the radius of gyration of p53 TAD (Supplementary Fig. 9), consistent with the observation in AA simulations⁴⁹. However, the absolute binding affinity was not predicted very accurately in this case. Compared with the binding strength 8.9 μ M measured by surface plasmon resonance (SPR) assays, the prediction by Martini3-IDP was around 20-fold weaker (Supplementary Table 9). Given the importance of W23 and several hydrophobic residues in stabilizing the binding mode, this result might point to the need for a potential improvement of the tryptophan and other apolar side chain models in Martini 3.

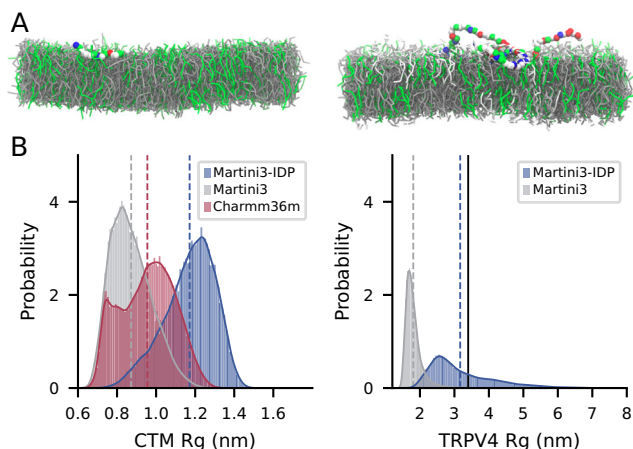


Fig. 4 | IDP-membrane interaction profile. A Last frame MD snapshots of Martini3-IDP simulations with CTM (left) binding to a lipid membrane consisting of POPC (70%, dark gray) and POPS (30%, green), TRPV4 (right) binding to a lipid membrane consisting of POPC (70%, dark gray) DOPS (20%, green) as well as PIP2 (10%, light gray). B Probability distribution of the radius of gyration of CTM/TRPV4 under solution condition, comparing Martini 3 and Martini3-IDP. Dotted lines represent the mean value of the distribution and the black solid line denotes the experimental Rg value reported for TRPV4⁵⁹.

Thirdly, we studied the binding between the Androgen receptor (AR) N-terminal transactivation domain (NTD) with the inhibitor EPI-002. This inhibitor has been pushed to clinical trials, and is one of the first examples of an IDP-binding small molecule to be tested in humans⁵⁵. To validate Martini3-IDP for this system, we performed 20 μ s simulations of AR-NTD with EPI-002. Reassuringly, the EPI-002 binding affinity was predicted in semi-quantitative agreement with the experimental measurements, suggesting a binding strength in the mM range (Supplementary Table 9). As for the Asyn system, dynamic shuttling of the ligand was observed, consistent with the behavior seen in AA simulations, including the preference for interacting with aromatic residues in the R2 and R3 regions (Fig. 3C).

Together, the tests presented in this section show the potential of Martini3-IDP in investigating the interaction between IDPs and small molecule ligands. We demonstrated the ability to obtain ligand absolute binding affinities and relative binding strength predictions with (in some cases) quantitative accuracy, possibly enabling candidate library screening and SAR analysis. In addition, the binding mode could also be captured well by Martini3-IDP, though resolving high resolution IDP-ligand complex conformation is still limited due to the lack of transient secondary structure description in the current Martini force field.

Martini3-IDP maintains IDP-lipid membrane binding

To make sure our developed Martini3-IDP model would not undermine the correct interaction profile with other classes of biomolecules in Martini, next, we verified the binding between IDPs and lipid membranes. Validating IDP-membrane binding is important for several reasons. First, from a biological perspective, it plays a key role in cell signaling and anchoring functions of IDP^{56,57}. Second, a key application area of Martini involves protein-membrane interactions that have been well characterized for Martini^{13,58,59}. And third, some recent temporary solutions for Martini 3 IDP conformational expansion are expected to affect the interaction profiles with other biomolecules. For instance, the protein-water rescaling strategy resulted in weaker interactions of IDPs with lipid membranes due to hydration competition¹⁵.

For validation purposes, we selected two IDPs which are known to interact with membranes, CTM (C-terminal motif of Complexin) and TRPV4 (Transient receptor potential cation channel subfamily vanilloid member 4). Both CTM and TRPV4 have been extensively

characterized by previous Martini simulations as well as experimentally, but their appropriate binding propensity could not be captured with the protein-water rescaling strategy^{18,59}. Here, we performed triplicate simulations of 3 and 5 μ s for CTM and TRPV4, respectively, interacting with different lipid bilayers (POPC:POPS = 7:3 and POPC:DOPS:PIP2 = 7:2:1 for CTM and TRPV4, respectively) following the same settings as in previous work (see Methods)¹⁵. As shown in Fig. 4, despite the expected IDP conformational expansion in Martini3-IDP compared to standard Martini 3 (bringing Martini3-IDP in principle closer to the experimental data for TRPV4, but slightly over-expanded for CTM) under solution conditions⁵⁹, their interaction propensity with the lipid membrane remained intact. This is also evident from the time evolution of the minimal distance between any beads of the IDPs and lipid membrane, indicating strong binding (Supplementary Fig. 10) in line with recently reported results using the same simulation settings¹⁵.

Based on this data, we conclude that the improvement of bonded parameters does not jeopardize the membrane affinity of known membrane-binding IDPs such as CTM and TRPV4. While the improved IDP conformational behavior in water is promising, further studies are necessary to validate whether conformational ensembles are accurately captured in membrane environments. More studies will be crucial to confirm Martini3-IDP's capability to correctly capture membrane binding modes across a wide range of systems, thereby enabling more accurate investigations of cell signaling processes involving IDPs or disordered protein regions in general.

LLPS in the Martini3-IDP model

Recently, formation of biomolecular condensates has attracted increasing interest, being implicated in a wide range of cellular functions, including local organization, gene expression regulation and misfolding protection. Biomolecular condensates can form via liquid-liquid phase separation (LLPS), mainly attributed to the presence of disordered and low-complexity domains (LCD) of proteins^{60–64}. Martini 3 has already been successfully used to model a variety of condensates^{65–68}. However, Martini 3 has a tendency toward too strong phase separation, underestimating the hydration level of the condensates which sometimes leads to formation of solid-like aggregates rather than coexisting liquid phases. To alleviate this problem, in recent work on LLPS simulations of a LCD protein such as TDP43, the protein-water rescaling strategy was adopted^{68,69}. We hypothesize the underlying problem is a direct consequence of the too compact IDP conformations in Martini 3: compact conformations reduce the conformation entropy loss in phase demixing and make the phase separation more thermodynamically favorable⁷⁰. To verify this hypothesis, a large diversity of LLPS systems were tested with the Martini3-IDP force field including homotypic, heterotypic, multiphase and MDP-containing systems.

Based on the final snapshots obtained after 16–32 μ s simulations (Fig. 5A, Supplementary Table 5), the typical homotypic LLPS systems LCD of RNA binding proteins hnRNP A1 (AILCD), protein Fused in Sarcoma (FUSLCD), GY23 peptide, mussel foot protein mfp-3S, and TIA1LCD, all showed stable phase separation. The corresponding time-dependent density profiles (Supplementary Fig. 11) confirm the stability of the bi-phasic nature of these systems. From the average density profiles (Fig. 5A), we can estimate the protein and water concentration in the dense and dilute phases. The dense phase protein concentration, around 500–700 mg/ml, is close to the experimental measurements of typical protein condensates. For instance, the dense phase density of FUSLCD is ~550 mg/ml, only slightly denser than 477 mg/ml protein concentration measured by NMR⁷¹. The dilute phase concentration, however, could only be measured for the short model peptide GY23 and was found to be ~14 mg/ml (~6 mM), around 10-fold higher than the experimental results showing that phase separation occurs at 0.5 mM or higher⁷². For comparison (Supplementary Fig. 12), the IDPs with more compact conformation in standard Martini 3 indeed showed

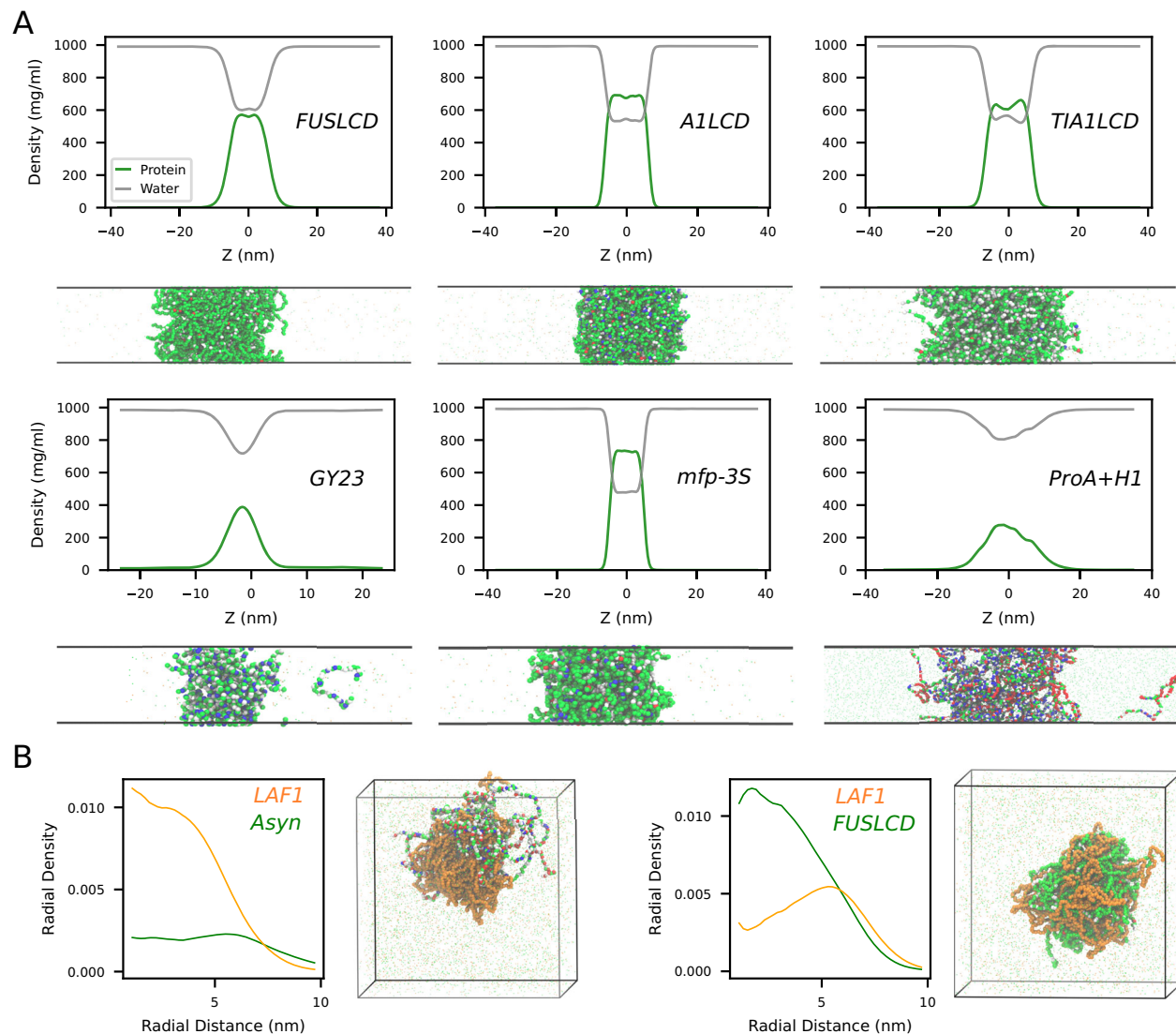


Fig. 5 | LLPS performance of Martini3-IDP. A Density profiles along the Z-axis, together with a snapshot of the last frame of the trajectory, for a variety of homotypic and heterotypic systems. Residues are color-coded white for apolar, green for uncharged polar, red for anionic and blue for cationic. Density profiles were averaged over the whole trajectory after discarding the initial equilibrium phase, and centered at the center-of-mass of the condensate ($Z=0$). **B** Snapshots and radial bead density profiles of multiphase condensates. Orange chains represent LAF1 and residue type specific chains represent Asyn and FUSLCD (Residues are color-coded

white for apolar, green for uncharged polar, red for anionic and blue for cationic). Density profiles were obtained with respect to the center-of-geometry of the condensate, orange lines represent LAF1, green lines represent Asyn or FUSLCD. FUSLCD, A1LCD, TIA1LCD represent the LCD of RNA binding proteins hnRNPA1, protein Fused in Sarcoma and TIA-1. GY23 represents GY23 peptide, mfp-3S represents mussel foot protein mfp-3S. ProA+H1 represents the mixture of Histone H1 and its nuclear chaperone Prothymosin-a. LAF1 represents the N-terminal intrinsically disordered RGG region of ATP-dependent RNA helicase LAF-1.

relatively stronger LLPS propensity and denser condensates, deviating from the experimental water content more. Finally, to prove that the dense phase is a liquid and not a solid aggregate, we characterized the dynamics of these systems through calculation of the mean square displacement (MSD) of protein and water inside the bulk and condensate regions (Supplementary Table 7 and Supplementary Fig. 13). This data reveals protein diffusion coefficients of the order of $10^{-11}\text{m}^2\text{s}^{-1}$ to $10^{-10}\text{m}^2\text{s}^{-1}$ inside the condensates, corresponding to liquid-like dynamics and close to the values reported in previous research^{65,71}. Besides, the slowdown of water dynamics from bulk water to condensate phase is well captured.

Heterotypic condensate formation by two highly and oppositely charged disordered proteins, Histone H1 and its nuclear chaperone Prothymosin-a (ProA) was also investigated. Note that H1 is a multi-domain protein and the full-length protein was simulated. Based on a 24 μs simulation, coexisting liquid phases with dynamic protein

monomer exchange could be observed (Fig. 5A). The resulting dilute/dense phase protein concentration of $\sim 1.8/280\text{ mg/ml}$ is quantitatively consistent with the experimental observations in $>150\text{ mM}$ salt condition (~ 1 and 220 mg/ml protein concentration)⁷³, although it should be noted that we used a higher salt concentration (600 mM) to allow sufficient screening given the relatively high protein concentration used in the simulations.

In realistic cellular environments, multiple discrete membraneless organelles are present in which multiple components coexist. The wide interactome of IDPs mediated by multivalent interactions poses challenges for the simulation of multi-component condensates, requiring high-quality description of both homotypic and heterotypic interactions. To test whether or not Martini3-IDP is up for this challenge, we consider mixtures of the N-terminal intrinsically disordered region of ATP-dependent RNA helicase LAF-1 (LAF1 RGG) with either Asyn or FUSLCD, which were reported to be immiscible⁷⁴. As visible

from the snapshots at the end of a 10 μ s simulation (Fig. 5B), both systems form a heterogeneous condensate. However, the internal organization of the condensate droplets differs between these two systems. In case of LAF1-Asyn, LAF1 proteins occupy the interior of the condensate with Asyn proteins being adsorbed and enriched on the surface. On the contrary, in the LAF1-FUSLCD mixture, the FUSLCD becomes the dominant nucleus component whereas LAF1 forms a condensate on the surface, although partial miscibility was observed as is also evident from the radial density distribution profiles (Fig. 5B). This behavior is consistent with previous simulations based on a one-bead per residue model, and is also observed in *in vivo* fluorescence experiments⁷⁴. The ability to segregate different components into discrete phases inside the condensates provides more evidence for the well-balanced protein interactions in Martini3-IDP, and paves the way for future realistic simulations of membraneless organelles in a cellular environment.

Discussion

Too compact disordered protein conformations observed in Martini 3, one of the most widely used biomolecular CG force fields, limits its applications. Some recent work proposed to rescale protein-water or protein-protein interactions in Martini 3 to expand protein conformations^{12,14,18}, following trends in atomistic simulation^{19,23}. However, this solution is not transferable for a number of reasons. Firstly, there are no clear signs that Martini 3 protein interactions are generally too strong to cause conformation collapse, and actually most recently tests claimed Martini 3 proteins are less sticky after alleviating the over-aggregation problem in Martini 2^{20–22}. Secondly, even if protein interaction propensity is too strong, the uniform and prefixed rescaling factor would not be optimal for all sequences. Some realistic local contacts may be disturbed while some collapse still exists under the uniform rescaling factor. Finally, it may disturb the intermolecular interactions with other biomolecules. For example, protein-water rescaling has already been shown to undermine IDP-lipid membrane binding due to the preference of IDPs to reside in the aqueous solution; and rescaling protein-protein interactions by a factor of 0.88 results in a complete loss of transmembrane protein self-association¹⁸ and LLPS of the FUSLCD (Supplementary Fig. 20).

In this study, by comparing conformational distributions sampled with Martini 3 to those obtained with an AA force field, we found that there was a huge discrepancy between the preferred Martini 3 and atomistic configurations. This prompted us to optimize the bonded parameter definition for Martini IDPs, resulting in the Martini3-IDP model. The major optimizations of bonded terms are summarized here:

1. The equilibrium BB-BB bond length and BB-BB-BB angle were slightly adjusted to better fit the atomistic reference distributions. In addition, Pro specific bond lengths were redefined due to the unique mapping of Proline in Martini protein models.
2. A generic BB-BB-BB-BB backbone dihedral was defined to replace the totally free sampling in Martini, using a combination of multiple proper dihedrals (Gromacs type 9 proper dihedral function), in line with recent observations of our group¹⁷ that disordered protein backbones have preferred orientations rather than being random-like.
3. BB-BB(-1)-BB(+1)-SC1 improper dihedrals, deduced from the atomistic reference configurations, were introduced to bias the side chain orientation relative to the connected backbone beads using a Gromacs type 2 harmonic improper dihedral function.
4. Auxiliary, weak SC1-BB-BB-SC1 dihedrals were defined for optimized fitting to the atomistic reference data, with specific consideration of Gly-containing sequences (BB-BB(Gly)-BB-SC1 dihedral potential), using Gromacs type 9 proper dihedral functions.
5. BB-BB-SC1 and SC1-BB-BB equilibrium angles were optimized by residue-specific fitting of the reference distributions.

Of note, the newly introduced BB-BB(-1)-BB(+1)-SC1 improper dihedral reflects the atomistic Ca *sp*³-hybrid tetrahedral configuration and is therefore a protein intrinsic feature. Inclusion of this term could help Martini to generate better configurations by coupling to the existing parameters. For instance, the neighboring side chain orientation, SC1-BB-BB-SC1 and broader SC1-BB-BB-BB/BB-BB-BB-SC1 dihedral distributions could be captured just by the coupling of the BB-BB(-1)-BB(+1)-SC1 improper dihedral and BB-BB-BB-BB dihedral. Moreover, the same concept could also be applied, in principle, to generate the outward side chain orientation in helical motifs and anti-parallel neighboring side chain orientation in sheet-like structures, via coupling of the backbone BB-BB-BB-BB dihedral (-60°/180° for helix/sheet) and local BB-BB(-1)-BB(+1)-SC1 improper dihedral. Therefore, this improper dihedral not only played a role in the optimization of disordered proteins in Martini 3, but could also be useful to improve other secondary structure elements, helping to avoid the use of side chain corrections which are dependent on a reference initial configuration.

The Martini 3 disordered protein model, Martini3-IDP, featuring these optimized bonded parameters, achieves a better compromise between accurate conformational description and improved radius of gyration, with quantitative accuracy compared to experimental SAXS data over a wide range of IDPs. It reinforces the key role of the protein-bonded parameter in disordered protein sampling. And compared with other Martini IDP ad hoc solutions, this model was optimized only based on fitting with atomistic resolution configurations, keeping the level of transferability as high as possible. Hence, the advantage of Martini3-IDP is not limited to IDP Rg prediction, but can be naturally integrated with the wide chemical space covered by current Martini models, thereby showing an advantage over other specifically trained disordered protein CG force fields^{25–27}. To validate it, a diverse range of test systems was considered, including multidomain proteins, IDP-ligand binding, IDP-lipid membrane binding and LLPS.

Like any model, Martini3-IDP also has some limitations: (1) Protein secondary structure transition is still not allowed in the current Martini protein framework, so Martini3-IDP does not sample the transient formation of helices or beta-sheets that are known to exist in certain IDPs; (2) Although Martini3-IDP already highly reduces the simulation costs compared to atomistic force fields as an alternative solution, the large size system simulations with explicit solvent still demand significant resources, such as the conformational ensemble sampling of the multidomain TDP43 protein with more than 400 residues length as well as the LLPS systems, where Martini sampling benefits from coupling to one-bead-per-residue models⁶⁷; (3) This Martini3-IDP version is built on top of the Martini 3 non-bonded parameters, whereas, preliminary benchmark tests indicate that Martini 3 side chain-side chain interactions leave room for improvement, and there is ongoing effort to optimize these side chains; (4) Not all residue sequences were sufficiently present in our reference database to consider all possible residue specific effects. In particular proline-rich sequences could benefit from further optimization of the bonded potentials.

Concluding, this study identified the lack of optimized bonded parameters being the main reason for Martini 3 IDP conformational collapse. The Martini3-IDP model captures the reference distributions obtained with an AA force field well, resulting in much better general agreement with the experimentally measured radii of gyration. Martini3-IDP keeps the right balance in interactions of IDPs with other biomolecules, enabling the Martini community to investigate a broad range of cellular processes involving IDPs.

Methods

IDP simulation setup

A set containing nine diverse IDPs was used for first-stage evaluation of standard Martini 3 performance of IDP conformations. A more

comprehensive set containing 22 IDPs was used for evaluation of the optimized Martini3-IDP model. IDP topology and initial structure were generated by Polyply⁷⁵ based on sequence fasta files. In Polyply generation, the default coil secondary structure was used for disordered proteins. IDPs were inserted into a dodecahedron box with system-dependent box size, avoiding periodic interference. Salt concentration was set to match the conditions in SAXS experiments and kept charge neutral. Details of the system setup of each IDP system are provided in Supplementary Table 1,2. Radius of gyration was computed using *gmx gyrate*. Each IDP was simulated between 10 and 25 μ s, ensuring simulation convergence (see Supplementary Fig. 14). Simulation convergence and error estimation has been quantified by performing blocking analysis (*gmx analyze -ee*), see Supplementary Fig. 19. For some IDP systems with multiple experimental Rg values reported, to avoid selection bias, the mean value of these Rg values were used. Error bars on the experimental Rg values were determined by a fitting error from the Guinier fit plus a standard deviation over reported experimental values from the literature. KL divergences between bonded parameters were computed by Scipy, provided in Supplementary Table 10.

Multidomain protein simulation setup

Similar to the setup for IDPs, the IDRs identified in multidomain proteins were generated by Polyply⁷⁵, based on sequence fasta files. The folded domains were coarse-grained to Martini by Martinize2⁷⁶ and reference AA structures were taken from the RCSDB database, predicted by AlphaFold⁷⁷ or modeled by Swiss-Model^{78,79}. An intradomain elastic network model was applied with Martinize2 to keep the folded domain intact, using harmonic potentials of 700 kJ mol⁻¹ nm⁻² between backbone beads within 0.9 nm cutoff. Secondary structure was assigned with DSSP⁸⁰ in Martinize2. The *-scFix* flag was used to restrict side chain orientation in the folded domain based on reference AA structure²⁶.

Topology files of IDRs and folded domains were merged together by treating the linking regions as disordered. Generic sequence-independent backbone bond length, BB-BB-BB angle, BB-BB-BB-BB dihedral, SC1-BB-BB-SC1 dihedral and SC1-BB-BB/BB-BB-SC1 angle from Martini3-IDP were defined for linking regions. In addition, the terminal protonation state in each fragment was corrected manually after merge. A simple script for topology merge is provided on Github (<https://github.com/Martini-Force-Field-Initiative/Martini3-IDP-parameters>); Similarly, the Martini3-IDP force field was also implemented in Martinize2 to easily generate multidomain protein topology generation in one step, only providing initial AA structure and defining the IDR region. The Martinize2 format force field file is also provided on Github (<https://github.com/Martini-Force-Field-Initiative/Martini3-IDP-parameters>). The full length multidomain initial structure at Martini resolution could be coarse grained by Martinize2, and the reference AA structure was predicted by AlphaFold or generated by merging the fragment structures together after a short simulation.

Each MDP was inserted into a dodecahedron box with system-dependent box size, keeping protein concentration and salt concentration to match the conditions in SAXS experiments. Setting details of MDP systems for Rg evaluation are listed in Supplementary Table 3. Each MDP was simulated between 20 and 30 μ s, ensuring simulation convergence (see Supplementary Fig. 15). Simulation convergence and Rg error estimation was quantified by performing blocking analysis (*gmx analyze -ee*), shown in Supplementary Fig. 18. Dissociation constant and error in the CTPR simulation was estimated by blocking analysis in each of the parallel replicas with different temperatures. Dissociation constant and error in the CXCL-HMGB1 complex simulation was estimated from statistics of three replicas.

IDP-small molecule ligand binding setup

Asyn, p53 TAD and AR NTD topology files were generated as described above for individual IDPs. Small molecule parameterization is

described in more detail in the SI, final topologies are available on Github and from the Martini Database (MAD)⁸¹. Asyn-ligand systems were simulated in a cubic water box with 10.8 nm sides, ions were added to a concentration of 25 mM for Ligands 23 and 47, and to a concentration of 50 mM for Fasudil, consistent with the reference work from D.E Shaw research⁵⁴. p53 TAD and AR NTD ligand systems were simulated in dodecahedron water boxes with 13.5 and 18 nm sides, at 150 and 20 mM salt concentration, respectively, consistent with the reference work^{49,55}. Production simulation of each system was performed for 20-50 μ s, sufficient to observe enough association-dissociation events for obtaining accurate binding statistics. Full details of the system setup are provided in Supplementary Table 4.

To analyze the IDP-ligand binding, the minimum distance and contact number between different components were calculated by *gmx mindist*. The sequence-specific binding frequency was calculated using the MDAnalysis python package^{82,83}. Bound states were defined as configurations for which the minimum distance between any ligand and residue bead was smaller than 0.6 nm. The contact profiles along the trajectory are provided in Supplementary Figs. 16 and 17. The binding affinity (dissociation constant K_D) was calculated from⁸⁴:

$$K_D = \frac{P_u}{P_b} (\nu N_{av})^{-1} \quad (1)$$

where P_u is the fraction of simulation time in which the protein and ligand are unbound, P_b is the fraction of simulation time in which the protein and ligand are bound, ν is the volume of the simulation box, and N_{av} is Avogadro's number. Error estimation was calculated from blocking analysis.

IDP-lipid membrane binding simulation setup

Lipid bilayers in this section were generated by the Insane script⁸⁵ with 20 × 20 nm lateral dimension. The composition of the membrane, POPC:POPS = 7:3 and POPC:DOPS:PIP2 = 7:2:1 for CTM and TRPV4, respectively, was chosen to be consistent with the reference work¹⁵. To simulate protein-membrane interactions, initial structures were generated with a minimum distance of 3 nm between any bead of protein and any bead of the lipids, resulting in a Z-axis length of 13 and 23 nm for CTM and TRPV4, respectively. 150 mM NaCl ions were added. Three replicate simulations of each system were performed for 3 (CTM) and 5 (TRPV4) μ s, sufficient to observe membrane binding. Membrane binding was quantified based on defining bound states when the minimum distance between any beads of the protein and lipids was smaller than or equal to 0.7 nm¹⁵. The topology generation for the CTM and TRPV4 IDPs was the same as described above for IDP simulations (setup details of CTM and TRPV4 monomers are also listed in Supplementary Table 2).

A 2 μ s reference CHARMM36m atomistic simulation of the CTM system was performed for comparison of CTM conformational behavior in the solution environment, as in Fig. 4B. Setup details of this CTM system are given in Supplementary Table 6.

LLPS simulation setup

Direct-coexistence simulations for homotypic and heterotypic LLPS were performed in a rectangular box⁸⁶. The X/Y axis dimensions were set to 8-14 nm, with the Z-dimension 50 nm or more, as listed in Supplementary Table 5. Based on finite-size analysis, 8 nm lateral dimension was enough to obtain robust coexisting phases for FUSLCD modeled with the Mpipi force field²⁰. Depending on the protein, between 30–60 copies of the proteins were included to keep the dense phase thick enough and simulation tractable. The protein copies were inserted into a relatively small slab region as initial configuration to accelerate LLPS and simulation convergence. The simulation box was then expanded to 50-80 nm along the Z axis. A salt concentration of 150/100 mM was used according to experimental conditions (see

Supplementary Table 5). For the ProA+H1 system, a higher salt concentration (600 mM KCl) was used in this simulation to mimic the experimental electrostatic screening considering the high charge content in this system (+53 net charge on H1 and -44 net charge on ProA) and relatively high protein concentration used in MD simulations; previously it was shown that, in case of a pHis/pGlu complex coacervate, an increase of 2 to 10 times the polymer concentration promotes a shift of the transition point to 0.2–0.5 M higher salt concentration^{65,87}. For all systems, between 16–32 μ s production simulations were run, long enough to make sure the density profiles remained unchanged when evaluated over successive time windows (Supplementary Fig. 11).

Multiphase LLPS simulations were performed in a cubic box with 28 nm edges, for mixtures of 15 LAF1 RGG:15 FUSLCD and 15 LAF1 RGG:12 Asyn. To ensure formation of a droplet, the proteins were first inserted randomly inside a slightly smaller box of 23 nm edge, then expanded to 28 nm, and subsequently solvated. A relatively high concentration of 300 mM salt was added to increase the electrostatic screening of the high charge content of these proteins. This was followed by a 6 μ s production run to observe droplet formation and reach a stable internal organization in the last 500 ns. Setup details of all LLPS systems can be found in Supplementary Table 5.

Mass density profiles (Supplementary Fig. 11) were determined along the Z-axis with protein density centered by the *gmx density* tool. Considering bead masses in Martini 3 are usually slightly larger than actual atom group masses, a rescaling factor (the ratio of protein molar weight and beads molar weight in Martini) was used to correct the mass density profile. If phase separation occurs, the protein concentration of the dilute and concentrated phases was obtained as the average concentration in the corresponding region. Radial density profiles were computed by calculating protein bead density in each radial distance bin relative to the condensate center with bin width equal to 0.2 nm, using python and the MDAnalysis package^{82,83}.

Diffusion coefficients were calculated from the mean-squared-displacement (MSD) by *gmx msd* using the Einstein relationship. For the MSD, protein or water beads residing in small volume elements well inside the dilute or condensate phase were selected to avoid an influence of the interfacial region. We limited the fitting of the MSD to the range of times 0 to 200 ps, where there are no deviations from linearity.

General simulation protocol

In the CG simulations, following the standard protocol¹¹, the Verlet scheme was used with the default energy drift buffer. The Verlet neighbor search algorithm is used to update the neighbor list every 20 steps with a buffer tolerance of 0.005 kJ mol⁻¹ps⁻¹. By default, nrexcl = 1 setting is applied in Martini, implying 1–4 non-bonded interactions are on. Potential shift was used to calculate van der Waals interactions with 1.1 nm cutoff and Coulomb interactions were treated by a reaction-field with dielectric constant 15 and 1.1 nm cutoff. Production simulations were run with 20 fs step size, using velocity-rescaling thermostat (temperature coupling constant 1 ps) and Parrinello-Rahman barostat (pressure coupling constant 12 ps). For membrane systems, semi-isotropic pressure coupling was implemented independently in the XY plane and Z direction. In direct-coexistence LLPS simulations, semi-isotropic pressure coupling was used keeping the XY plane uncompressed. Isotropic pressure coupling was used in all other systems. All simulations were executed by Gromacs, versions 2021.5, 2021.7, and 2022.6⁸⁸.

AA level IDP simulation setup

Microsecond-scale CHARMM36m atomistic trajectories of 12 IDPs, part of the recent Martini 3 benchmark, served as reference data^{14,17}. To augment the AA reference set, four additional short IDPs (RS-peptide, FG-peptide, HEWL19 and polyQ30) were simulated in this study for 1–1.5 μ s, also using the CHARMM36m force field²⁷. Individual proteins

were inserted in a dodecahedron box, solvated with 150 mM NaCl (except for polyQ30 for which no salt was added). The Verlet-scheme was used with the default energy drift buffer. Cutoff for van der Waals interactions was set to 1.2 nm with a 1.0–1.2 nm switch. The smooth particle mesh Ewald (PME) was used for Coulomb interactions, with a real-space cutoff of 1.2 nm. The CHARMM-modified TIP3P water model was used as recommended²⁷. AA resolution trajectories were mapped to Martini resolution using fast-forward⁸⁹. Setup details of the 4 AA IDP systems are given in Supplementary Table 6.

Principal component analysis

PCA based on the pairwise distances between alpha-Carbons was used to project conformational ensembles from different force fields. Martini trajectories were back-mapped to AA structures using the Backward algorithm⁹⁰. For each protein, all ensembles were pooled together for PCA in order to project onto the same two main principal components. PCA was performed with PyEMMA⁹¹. For a quantitative estimation of the conformational space sampling in each force field, the conformational probability distribution in 2D PCA space was discretized into bins with a binwidth of 10, then the total Shannon entropy was measured along all bins by Scipy⁹². The Amber99SB-disp force field simulation of 3 IDPs were taken from Robustelli et al.³¹. The NMR-derived ensembles were extracted from the PED dataset^{93,94}.

Optimization strategy

Considering the multiple bonded terms in our model (high-dimensional parameter space), complex couplings between different bonded terms and the resource cost of Martini-resolution IDP simulation, no quantitative objective function was used in our work to iteratively and automatically optimize the parameters of bonded terms. Instead, the bonded potentials were manually adjusted targeting two major objectives, namely matching the bonded distributions from our reference AA simulations and providing a better overall match to the experimentally measured radii of gyration across the entire database.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Force field files for Martini3-IDP and related data generated in this paper have been deposited in Github repository under accession <https://github.com/Martini-Force-Field-Initiative/Martini3-IDP-parameters> and a backup of this repository is available via <https://doi.org/10.5281/zenodo.14968191>. Structure data for most simulations presented in this work have been deposited in Zenodo under accession <https://zenodo.org/records/14608855>. Source data are provided with this paper.

Code availability

Scripts and code used in this paper have been deposited in Github under accession <https://github.com/Martini-Force-Field-Initiative/Martini3-IDP-parameters> and a backup of this repository is available via <https://doi.org/10.5281/zenodo.14968191>.

References

- Panca, R. & Tompa, P. Structural disorder in eukaryotes. *PLoS ONE* **7**, e34687 (2012).
- Xue, B., Dunker, A. K. & Uversky, V. N. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* **30**, 137–149 (2012).
- Holehouse, A. S. & Kragelund, B. B. The molecular basis for cellular function of intrinsically disordered protein regions. *Nat. Rev. Mol. Cell Biol.* **25**, 187–211 (2024).

4. Brodsky, S., Jana, T. & Barkai, N. Order through disorder: The role of intrinsically disordered regions in transcription factor binding specificity. *Curr. Opin. Struct. Biol.* **71**, 110–115 (2021).
5. Fuxreiter, M. et al. Malleable machines take shape in eukaryotic transcriptional regulation. *Nat. Chem. Biol.* **4**, 728–737 (2008).
6. Zhang, X., Bai, X. & Chen, Z. J. Structures and mechanisms in the cGAS-STING innate immunity pathway. *Immunity* **53**, 43–53 (2020).
7. Cuylen, S. et al. Ki-67 acts as a biological surfactant to disperse mitotic chromosomes. *Nature* **535**, 308–312 (2016).
8. Witus, S. R. et al. BRCA1 / BARD1 intrinsically disordered regions facilitate chromatin recruitment and ubiquitylation. *EMBO J.* **42**, e113565 (2023).
9. Lyon, A. S., Peeples, W. B. & Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **22**, 215–235 (2021).
10. Schmidt, H. B. & Görlich, D. Transport selectivity of nuclear pores, phase separation, and membraneless organelles. *Trends Biochem. Sci.* **41**, 46–61 (2016).
11. Borges-Araújo, L. et al. Pragmatic coarse-graining of proteins: models and applications. *J. Chem. Theory Comput.* **19**, 7112–7135 (2023).
12. Souza, P. C. T. et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **18**, 382–388 (2021).
13. Marrink, S. J. et al. Two decades of Martini: Better beads, broader scope. *WIREs Comput. Mol. Sci.* **13**, e1620 (2023).
14. Thomasen, F. E., Pesce, F., Roesgaard, M. A., Tesei, G. & Lindorff-Larsen, K. Improving Martini 3 for disordered and multidomain proteins. *J. Chem. Theory Comput.* **18**, 2033–2041 (2022).
15. Larsen, A. H. et al. Combining molecular dynamics simulations with small-angle X-ray and neutron scattering data to study multi-domain proteins in solution. *PLOS Comput. Biol.* **16**, e1007870 (2020).
16. Borges-Araújo, L., Pereira, G. P., Valério, M. & Souza, P. C. T. Assessing the Martini 3 protein model: A review of its path and potential. *Biochim. Biophys. Acta BBA Proteins Proteom.* **1872**, 141014 (2024).
17. Souza, P. C. T. et al. GōMartini 3: From large conformational changes in proteins to environmental bias corrections. Preprint at <https://doi.org/10.1101/2024.04.15.589479> (2024).
18. Thomasen, F. E. et al. Rescaling protein-protein interactions improves Martini 3 for flexible proteins in solution. *Nat. Commun.* **15**, 6645 (2024).
19. Best, R. B., Zheng, W. & Mittal, J. Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* **10**, 5113–5124 (2014).
20. Lamprakis, C. et al. Evaluating the efficiency of the martini force field to study protein dimerization in aqueous and membrane environments. *J. Chem. Theory Comput.* **17**, 3088–3102 (2021).
21. Claveras Cabezedo, A., Athanasiou, C., Tsengenes, A. & Wade, R. C. Scaling protein–water interactions in the Martini 3 coarse-grained force field to simulate transmembrane helix dimers in different lipid environments. *J. Chem. Theory Comput.* **19**, 2109–2119 (2023).
22. Spinti, J. K., Neiva Nunes, F. & Melo, M. N. Room for improvement in the initial martini 3 parameterization of peptide interactions. *Chem. Phys. Lett.* **819**, 140436 (2023).
23. Mu, J., Liu, H., Zhang, J., Luo, R. & Chen, H.-F. Recent force field strategies for intrinsically disordered proteins. *J. Chem. Inf. Model.* **61**, 1037–1047 (2021).
24. Zerze, G. H., Zheng, W., Best, R. B. & Mittal, J. Evolution of all-atom protein force fields to improve local and global properties. *J. Phys. Chem. Lett.* **10**, 2227–2234 (2019).
25. Hu, Z., Sun, T., Chen, W., Nordenskiöld, L. & Lu, L. Refined bonded terms in coarse-grained models for intrinsically disordered proteins improve backbone conformations. *J. Phys. Chem. B* **128**, 6492–6508 (2024).
26. Joseph, J. A. et al. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nat. Comput. Sci.* **1**, 732–743 (2021).
27. Tesei, G., Schulze, T. K., Crehuet, R. & Lindorff-Larsen, K. Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc. Natl. Acad. Sci. USA* **118**, e2111696118 (2021).
28. Herzog, F. A., Braun, L., Schoen, I. & Vogel, V. Improved side chain dynamics in MARTINI simulations of protein–lipid interfaces. *J. Chem. Theory Comput.* **12**, 2446–2458 (2016).
29. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
30. Dey, S., MacAinsh, M. & Zhou, H.-X. Sequence-dependent backbone dynamics of intrinsically disordered proteins. *J. Chem. Theory Comput.* **18**, 6310–6323 (2022).
31. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. USA* **115**, E4758–E4766 (2018).
32. Van Der Lee, R. et al. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **114**, 6589–6631 (2014).
33. Martin, E. W. et al. Interplay of folded domains and the disordered low-complexity domain in mediating hnRNP A1 phase separation. *Nucleic Acids Res.* **49**, 2931–2945 (2021).
34. Taneja, I. & Holehouse, A. S. Folded domain charge properties influence the conformational behavior of disordered tails. *Curr. Res. Struct. Biol.* **3**, 216–228 (2021).
35. Ripka, J. F., Perez-Riba, A., Chaturbudy, P. K. & Itzhaki, L. S. Testing the length limit of loop grafting in a helical repeat protein. *Curr. Res. Struct. Biol.* **3**, 30–40 (2021).
36. Gurumoorthy, V. et al. Disordered domain shifts the conformational ensemble of the folded regulatory domain of the multidomain oncoprotein c-Src. *Biomacromolecules* **24**, 714–723 (2023).
37. Staby, L. et al. Disorder in a two-domain neuronal Ca²⁺-binding protein regulates domain stability and dynamics using ligand mimicry. *Cell. Mol. Life Sci.* **78**, 2263–2278 (2021).
38. Moses, D. et al. Structural biases in disordered proteins are prevalent in the cell. *Nat. Struct. Mol. Biol.* <https://doi.org/10.1038/s41594-023-01148-8> (2024).
39. Mantonico, M. V. et al. The acidic intrinsically disordered region of the inflammatory mediator HMGB1 mediates fuzzy interactions with CXCL12. *Nat. Commun.* **15**, 1201 (2024).
40. Jussupow, A. et al. The dynamics of linear polyubiquitin. *Sci. Adv.* **6**, eabc3786 (2020).
41. Zhang, Y. et al. Determining structural ensembles of flexible multi-domain proteins using small-angle X-ray scattering and molecular dynamics simulations. *Protein Cell* **6**, 619–623 (2015).
42. Michie, K. A., Kwan, A. H., Tung, C.-S., Guss, J. M. & Trewthella, J. A highly conserved yet flexible linker is part of a polymorphic protein-binding domain in myosin-binding protein C. *Structure* **24**, 2000–2007 (2016).
43. Lin, Y.-H. et al. The intrinsically disordered N-terminal domain of galectin-3 dynamically mediates multisite self-association of the protein through fuzzy interactions. *J. Biol. Chem.* **292**, 17845–17856 (2017).
44. Wang, Y. et al. Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein. *Proc. Natl. Acad. Sci. USA* **110**, E3743–E3752 (2013).
45. Brodsky, S. et al. Intrinsically disordered regions direct transcription factor in vivo binding specificity. *Mol. Cell* **79**, 459–471.e4 (2020).
46. Borgia, A. et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61–66 (2018).

47. Banerjee, P. R. et al. Dissecting the biophysics and biology of intrinsically disordered proteins. *Trends Biochem. Sci.* **49**, 101–104 (2024).
48. Wang, H., Xiong, R. & Lai, L. Rational drug design targeting intrinsically disordered proteins. *WIREs Comput. Mol. Sci.* **13**, e1685 (2023).
49. Ruan, H. et al. Computational strategy for intrinsically disordered protein ligand design leads to the discovery of p53 transactivation domain I binding compounds that activate the p53 pathway. *Chem. Sci.* **12**, 3004–3016 (2021).
50. Patel, K. N., Chavda, D. & Manna, M. Molecular docking of intrinsically disordered proteins: challenges and strategies. In *Protein-Protein Docking*, Vol. 2780 (ed. Kaczor, A. A.) 165–201 (Springer US, New York, NY, 2024).
51. Diamanti, E. et al. Identification of inhibitors targeting the energy-coupling factor (ECF) transporters. *Commun. Biol.* **6**, 1182 (2023).
52. Souza, P. C. T. et al. Protein–ligand binding with the coarse-grained Martini model. *Nat. Commun.* **11**, 3714 (2020).
53. Kjølbye, L. R. et al. Towards design of drugs and delivery systems with the Martini coarse-grained model. *QRB Discov.* **3**, e19 (2022).
54. Robustelli, P. et al. Molecular basis of small-molecule binding to α -synuclein. *J. Am. Chem. Soc.* **144**, 2501–2510 (2022).
55. Zhu, J., Salvatella, X. & Robustelli, P. Small molecules targeting the disordered transactivation domain of the androgen receptor induce the formation of collapsed helical states. *Nat. Commun.* **13**, 6390 (2022).
56. Kjaergaard, M. & Kragelund, B. B. Functions of intrinsic disorder in transmembrane proteins. *Cell. Mol. Life Sci.* **74**, 3205–3224 (2017).
57. Mangiarotti, A. & Dimova, R. Biomolecular condensates in contact with membranes. *Annu. Rev. Biophys.* **53**, 319–341 (2024).
58. Srinivasan, S., Zoni, V. & Vanni, S. Estimating the accuracy of the MARTINI model towards the investigation of peripheral protein–membrane interactions. *Faraday Discuss* **232**, 131–148 (2021).
59. Goretzki, B. et al. Crosstalk between regulatory elements in disordered TRPV4 N-terminus modulates lipid-dependent channel activity. *Nat. Commun.* **14**, 4165 (2023).
60. Pappu, R. V., Cohen, S. R., Dar, F., Farag, M. & Kar, M. Phase transitions of associative biomacromolecules. *Chem. Rev.* **123**, 8945–8987 (2023).
61. Ripin, N. & Parker, R. Formation, function, and pathology of RNP granules. *Cell* **186**, 4737–4756 (2023).
62. Martin, E. W. & Holehouse, A. S. Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof. *Emerg. Top. Life Sci.* **4**, 307–329 (2020).
63. Protter, D. S. W. et al. Intrinsically disordered regions can contribute promiscuous interactions to RNP granule assembly. *Cell Rep.* **22**, 1401–1412 (2018).
64. Dignon, G. L., Best, R. B. & Mittal, J. Biomolecular phase separation: from molecular driving forces to macroscopic properties. *Annu. Rev. Phys. Chem.* **71**, 53–75 (2020).
65. Tsanai, M., Frederix, P. W. J. M., Schroer, C. F. E., Souza, P. C. T. & Marrink, S. J. Coacervate formation studied by explicit solvent coarse-grain molecular dynamics with the Martini model. *Chem. Sci.* **12**, 8521–8530 (2021).
66. Liu, Y., Wang, X., Wan, Z., Ngai, T. & Tse, Y.-L. S. Capturing coacervate formation and protein partition by molecular dynamics simulation. *Chem. Sci.* **14**, 1168–1175 (2023).
67. Brasnett, C., Kiani, A., Sami, S., Otto, S. & Marrink, S. J. Capturing chemical reactions inside biomolecular condensates with reactive Martini simulations. *Commun. Chem.* **7**, 151 (2024).
68. Ingólfsson, H. I. et al. Multiscale simulations reveal TDP-43 molecular-level interactions driving condensation. *Biophys. J.* **122**, 4370–4381 (2023).
69. Gaurav, K. et al. Multi-scale simulations reveal molecular drivers in MUT-16 scaffold protein phase separations and client recognition. Preprint at <https://doi.org/10.1101/2024.04.13.589337> (2024).
70. Li, L. & Hou, Z. Crosslink-induced conformation change of intrinsically disordered proteins have a nontrivial effect on phase separation dynamics and thermodynamics. *J. Phys. Chem. B* **127**, 5018–5026 (2023).
71. Murthy, A. C. et al. Molecular interactions underlying liquid–liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.* **26**, 637–648 (2019).
72. Gabryelczyk, B. et al. Hydrogen bond guidance and aromatic stacking drive liquid–liquid phase separation of intrinsically disordered histidine-rich peptides. *Nat. Commun.* **10**, 5465 (2019).
73. Galvanetto, N. et al. Extreme dynamics in a biomolecular condensate. *Nature* **619**, 876–883 (2023).
74. Welles, R. M. et al. Determinants that enable disordered protein assembly into discrete condensed phases. *Nat. Chem.* **16**, 1062–1072 (2024).
75. Grünewald, F. et al. Polyply; a python suite for facilitating simulations of macromolecules and nanomaterials. *Nat. Commun.* **13**, 68 (2022).
76. Kroon, P. et al. Martinize2 and Vermouth: unified framework for topology generation. *eLife*. **12**, RP90627 (2024).
77. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
78. Bienert, S. et al. The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.* **45**, D313–D319 (2017).
79. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
80. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
81. Hilpert, C. et al. Facilitating CG simulations with MAD: the Martini Database Server. *J. Chem. Inf. Model.* **63**, 702–710 (2023).
82. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).
83. Gowers, R. et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. 98–105. <https://doi.org/10.25080/Majora-629e541a-00e> (2016).
84. Jong, D. et al. Determining equilibrium constants for dimerization reactions from molecular dynamics simulations. *J. Comput. Chem.* **32**, 1919–1928 (2011).
85. Wassenaar, T. A., Ingólfsson, H. I., Böckmann, R. A., Tieleman, D. P. & Marrink, S. J. Computational lipidomics with *insane*: a versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theory Comput.* **11**, 2144–2155 (2015).
86. Dignon, G. L., Zheng, W., Kim, Y. C., Best, R. B. & Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLOS Comput. Biol.* **14**, e1005941 (2018).
87. Li, L. et al. Phase behavior and salt partitioning in polyelectrolyte complex coacervates. *Macromolecules* **51**, 2988–2995 (2018).
88. Abraham, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
89. fast-forward https://github.com/fgrunewald/fast_forward.
90. Wassenaar, T. A., Pluhackova, K., Böckmann, R. A., Marrink, S. J. & Tieleman, D. P. Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *J. Chem. Theory Comput.* **10**, 676–690 (2014).
91. Scherer, M. K. et al. PyEMMA 2: A software package for estimation, validation, and analysis of markov models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).

92. Pauli, V. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
93. Ghafouri, H. et al. PED in 2024: improving the community deposition of structural ensembles for intrinsically disordered proteins. *Nucleic Acids Res.* **52**, D536–D544 (2024).
94. Teixeira, J. M. C. et al. IDPConformerGenerator: a flexible software suite for sampling the conformational space of disordered protein states. *J. Phys. Chem. A* **126**, 5985–6003 (2022).
95. Quemener, E., Corvellec, M. SIDUS—the solution for extreme deduplication of an operating system. *Linux J.* <https://doi.org/10.5555/2555789.2555792>.

Acknowledgements

L.W. was supported by the China Scholarship Council. We acknowledge the use of computational resources from the Dutch National Super-computer Snellius. S.J.M. is supported by funding from the European Research Council (ERC) with the Advanced grant 101053661 “COMP-O-CELL”. L.B.A and P.C.T.S would like to thank the support of the French National Center for Scientific Research (CNRS) and the funding from research collaboration agreements with PharmCADD. L.B.A and P.C.T.S also acknowledge the support of the Centre Blaise Pascal’s IT test platform at ENS de Lyon (Lyon, France) for the computer facilities. The platform operates the SIDUS solution developed by Emmanuel Quemener⁹⁵.

Author contributions

L.W. and S.J.M. conceived the overall study. L.W. performed and analyzed all simulations under the supervision of S.J.M. C.B. helped discussing during the study and provided some reference atomistic simulations. L.W. wrote the first draft of the manuscript with input from S.J.M., P.C.T.S, L.B.A and C.B. All authors contributed to the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-58199-2>.

Correspondence and requests for materials should be addressed to Siewert J. Marrink.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025