



HAL
open science

The diversity of molecular interactions involving intrinsically disordered proteins: A molecular modeling perspective

Ilinka Clerc, Amin Sagar, Alessandro Barducci, Nathalie Sibille, Pau Bernadó, Juan Cortés

► **To cite this version:**

Ilinka Clerc, Amin Sagar, Alessandro Barducci, Nathalie Sibille, Pau Bernadó, et al.. The diversity of molecular interactions involving intrinsically disordered proteins: A molecular modeling perspective. Computational and Structural Biotechnology Journal, 2021, 19, pp.3817-3828. <10.1016/j.csbj.2021.06.031>. <hal-03281983>

HAL Id: hal-03281983

<https://laas.hal.science/hal-03281983v1>

Submitted on 8 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

The diversity of molecular interactions involving intrinsically disordered proteins: A molecular modeling perspective

Ilinka Clerc^a, Amin Sagar^b, Alessandro Barducci^b, Nathalie Sibille^b,
Pau Bernadó^{b,*}, Juan Cortés^{a,1,*}

^aLAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

^bCentre de Biochimie Structurale. INSERM, CNRS, Université de Montpellier, France

Abstract

Intrinsically Disordered Proteins and Regions (IDPs/IDRs) are key components of a multitude of biological processes. Conformational malleability enables IDPs/IDRs to perform very specialized functions that cannot be accomplished by globular proteins. The functional role for most of these proteins is related to the recognition of other biomolecules to regulate biological processes or as a part of signaling pathways. Depending on the extent of disorder, the number of interacting sites and the type of partner, very different architectures for the resulting assemblies are possible. More recently, molecular condensates with liquid-like properties composed of multiple copies of IDPs and nucleic acids have been proven to regulate key processes in eukaryotic cells. The structural and kinetic details of disordered biomolecular complexes are difficult to unveil experimentally due to their inherent conformational heterogeneity. Computational approaches, alone or in combination with experimental data, have emerged as unavoidable tools to understand the functional mechanisms of this elusive type of assemblies. The level of description used, all-atom or coarse-grained, strongly depends on the size of the molecular systems and on the timescale of the investigated mechanism. In this mini-review, we describe the most rele-

*Corresponding authors

Email addresses: pau.bernado@cbs.cnrs.fr, juan.cortes@laas.fr

¹Lead contact

vant architectures found for molecular interactions involving IDPs/IDRs and the computational strategies applied for their investigation.

Keywords: Intrinsically disordered proteins, molecular interactions, molecular dynamics simulations, conformational sampling, molecular recognition elements, liquid-liquid phase separation.

1. Introduction

In the last few decades, Intrinsically Disordered Proteins and Regions (IDPs/IDRs) have emerged as key actors in multiple fundamental biological processes [1, 2]. Due to the lack of permanent secondary and tertiary structure, IDPs/IDRs are highly malleable molecules adapted to perform specialized functions that complement those of their globular counterparts [3, 4]. Intrinsic disorder is abundant in eukaryotic proteomes, where it contributes to the cellular complexity by participating in the vast majority of signaling and regulation events [5, 6]. Their amino acid sequence, rich in charged and non-structuring residues [7, 8], and often displaying low complexity [9], determines their lack of permanent structure. These sequence features have been widely used in bioinformatics approaches to identify disorder and function from proteomics data [10, 11, 12]. While some proteins display disorder all along the sequence (IDPs), in other cases disorder is only present in specific segments of the sequence, which are named IDRs [3, 4]. IDRs can be placed between globular domains (linkers), restricting their relative distance and orientation, or at the N- or C-termini as disordered tails of folded domains [13]. These distinct disordered protein architectures define the types of the resulting assemblies occurring upon binding to the biological partners (see Figure 1).

From a functional perspective, most of these disordered segments act as interaction specialists [14]. Their plasticity enables highly specific recognition by adapting their bound conformation to the physicochemical nature of the partner surface. These interactions are normally performed via evolutionary-conserved short linear interaction motifs (SLiMs) inserted within the chain [15, 16]. SLiMs,

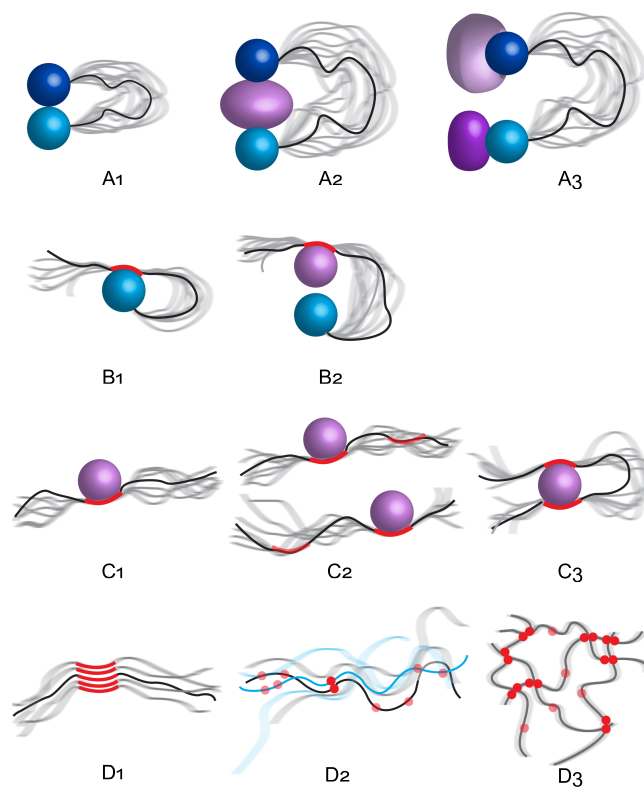


Figure 1: Illustration of different types of interactions involving IDPs/IDRs. They are classified depending on the ordered/disordered nature of the interacting regions. The representation is not aimed to be exhaustive, and combined interaction types do also exist. A-class cartoons refer to proteins consisting of multiple structured domains connected by flexible linkers. The domains can either interact intramolecularly (A₁) or with other biomolecules (A₂ and A₃). B-class cartoons represent interactions driven by SLiMs (in red) placed in IDRs that recognize their own globular domain (B₁) or another protein (B₂). C-class cartoons represent bimolecular interactions involving an IDP with a globular protein. While in C₁ assembly a single SLiM (in red) recognizes the globular domain, C₂ and C₃ represent scenarios where two similar SLiMs of the IDP interact with a globular protein with one (C₂) or two (C₃) binding sites. D-class cartoons represent the interaction between disordered proteins that either form amyloid-like structures (D₁), extremely fuzzy complexes (D₂) or unstructured condensates with liquid-like behavior (D₃). In D₂ and D₃, multiple low-affinity non-specific interactions (red dots) are present.

which encompass from 3 to 10 contiguous amino acids, are defined according to consensus sequences that are considered as the hot-spots of the interaction [16]. The large number and sequence variability of the identified interacting segments exemplify the richness of recognition events performed by IDRs [17]. Interestingly, several proteins share the same consensus sequence, identifying the family of binding partners recognized. Differences in the remaining residues and/or flanking regions can modulate the thermodynamics and kinetics of the recognition event, as well as the capacity to discriminate between several partners. Interaction mechanisms are classified according to the flexibility adopted by the disordered fragment upon binding [18]. While some disordered segments present a well-defined rigid structure in the bound form, others display an almost complete conformational freedom with multiple, very weak fuzzy contacts with the partner [18, 19, 20]. A prominent example of these extremely fuzzy complexes is the formation of liquid-like membrane-less compartments, which have appeared in the recent years as a very efficient mechanism for the spatio-temporal organization in living cells [21, 22]. Beyond the existence of these two extreme scenarios, the growing number of interactions reported suggests that there is a continuum of flexible binding modes [23]. Furthermore, post-translational modifications, often occurring in disordered segments modifying their physicochemical properties and enormously increasing the number of interaction possibilities [24], can act as switches to turn recognition events on and off [25, 26]. This large spectrum of interaction modes and regulation mechanisms explains the variability of functional outcomes and the numerous pathologies associated with the malfunction of disordered proteins and their complexes [27]. Note that although this review is focused on the interaction between IDPs/IDRs with other proteins, they can also be involved in interactions with small ligands, nucleic acids, lipids and carbohydrates [2, 28].

The conformational characterization of disordered proteins and their complexes still represents a challenge for biophysicists. The most suitable structural biology techniques for their study, Nuclear Magnetic Resonance (NMR) [29, 30], Small-Angle Scattering (SAS) [31, 32], single-molecule Förster Resonance En-

ergy Transfer (smFRET) [33], provide average information that reports on the ensemble of co-existing conformations present in solution [34, 35]. Moreover, interactions mediated by IDPs/IDRs are often characterized by their low affinity, inducing an equilibrium between bound and unbound forms that further complicates their structural characterization *in vitro* [36].

In this context, computational methods, either alone or in combination with experimental data, have become pivotal for the structural and dynamic characterization of this elusive class of biomolecules. A large variety of computational methods have been specifically developed, adapting the level of description to the size of the molecular system, the question to be addressed, and the availability of experimental information [37, 38, 39, 40, 41]. The final aim of most of these approaches is the generation of conformational ensembles representing realistic pictures of biomolecular entities with the capacity to provide the structural bases of cellular mechanisms and anticipate functional properties [42].

In this mini-review, we focus on how different computational strategies, alone or in combination with experimental data, have been applied to describe disordered biomolecular complexes. Note that the aim is not to provide an exhaustive enumeration of computational studies on IDPs/IDRs, but to briefly describe the methods and exemplify them with some relevant applications. After a succinct description of the various computational approaches usually applied to disordered complexes, we organized this mini-review according to the different architectures illustrated in Figure 1.

2. An overview of computational approaches

Various computational methods can be applied to the structural investigation of IDPs/IDRs and their interactions [37, 38, 39, 40, 41]. The choice of the method depends on different factors: (1) availability of experimental data, (2) level of detail and timescale at which the molecular mechanism has to be investigated, (3) size of the molecular system, (4) computing power available.

When experimental (biophysical or biological) information is lacking, bioin-

formatics tools can be applied to identify binding motifs from IDP sequences [43, 44], and to predict interactions between these motifs and protein partners [45, 46, 47]. These predictive tools deliver relevant insights to understand functional mechanisms involving IDPs. However, they only provide a partial and qualitative picture of molecular interactions. The study of thermodynamic and kinetic aspects of protein interactions requires a more global exploration of conformational states and transitions. This exploration can be based on different types of models and algorithms.

Whereas molecular dynamics (MD) simulations using all-atom physics-based force field models are widely used for the investigation of interactions involving globular proteins, the applicability of “standard” MD approaches to IDPs/IDRs is relatively limited [37, 48]. A first limitation comes from the fact that force fields, such as Amber and CHARMM, were mostly developed having globular proteins as targets. Thus, they tend to enrich the structure with secondary structure elements (α -helices and β -strands), and to produce collapsed conformations. Recent versions of these force-fields (e.g., [49, 50, 51, 52]) have been introduced to mitigate these effects. In particular, a balanced description of protein-water interactions thanks to new water models [53] or rescaling approaches [54] have been shown to be critical for improving the description of disordered proteins [51, 55].

The other limitation of all-atom MD simulations is their computational cost, which precludes their routine use to investigate large structural rearrangements or interaction mechanisms requiring long timescales. One of the main reasons for this high computational cost is the large size of the simulation box containing the protein(s) and water molecules, due to the large radius of gyration of IDPs (with respect to folded protein) and their fluctuations.

MD protocols applied to IDPs/IDRs often rely on enhanced sampling techniques, such as replica-exchange [56, 57], metadynamics [58, 59] or combined approaches [60, 61], which are more efficient than basic MD techniques to explore multiple-basin energy landscapes (we refer the interested reader to specialized reviews [62, 63] for further information on enhanced sampling techniques). Note

also that advances in software and hardware, enabling efficient parallel computing, have significantly contributed to extending the applicability of all-atom MD approaches, in particular thanks to the exploitation of graphics processing units (GPUs) [64]. Despite these methodological and technical advances, in practice, all-atom MD simulations are nowadays applicable to the investigation of relatively small systems (e.g. interactions involving protein fragments or a small number of disordered peptides) or too short timescales for larger systems.

The investigation of larger systems and/or longer timescales relies on the application of coarse-grained (CG) models. Although these models do not provide the same level of detail as the all-atom ones, they allow a much wider exploration of the conformational energy landscape. CG models can range from simple Gō-like models [65, 66] to more complex ones considering several beads per amino acid residue, such as AWSEM-IDP [67], PLUM [68], MARTINI [69] or SIRAH [70]. ABSINTH [71] can be considered as an intermediate approach between all-atom and CG models, since it only considers dihedral angles as variables, so that small groups of bonded atoms move as rigid bodies. For their application to IDPs, special attention has been paid to the (implicit) solvation terms included in most of these models. Note that implicit solvation models can be applied using other exploration algorithms, in addition to MD. This is for instance the case of ABSINTH, which was specially developed for Monte Carlo (MC) simulations [71].

Although MD-based methods are attractive due to their accuracy (particularly for atomistic simulations) and capacity to provide information on the temporal evolution of the molecular system, other types of algorithms are more efficient in sampling the huge conformational space of IDPs/IDRs. In addition to MC, several methods based on stochastic sampling techniques have been proposed to generate ensemble models of IDPs/IDRs. The most popular examples of these methods are TraDES [72] and Flexible-Meccano [73, 74]. These approaches incrementally construct IDP/IDR conformations using probability distributions of the dihedral ϕ and ψ angles of amino acid residues extracted from experimentally-determined protein structures, and can include informa-

tion about secondary structure propensities along the sequence. A recent variant of these methods, operating with three-residue fragments, has been shown to generate higher-quality conformational models of IDPs containing partially structured elements, which naturally emerge as they are encoded in the protein sequence [75].

While modeling approaches can provide an *ab initio* description of IDP/IDR conformational ensembles based only on physics- and/or knowledge-based models, their predictive capabilities can be greatly improved by taking advantage of available experimental information. In this respect, NMR, SAS, smFRET and other experimental results can be used for correcting the model inaccuracies, either by biasing or restraining the sampling into the most relevant regions of the conformational space, or by reweighting the simulation results *a posteriori*. Numerous algorithms have been proposed for this combination of simulation methods and experimental data (e.g. [76, 77, 78, 79, 80, 81, 82, 83, 84]). The interest to consider experimental data is particularly true for fast, stochastic approaches to generate conformational ensemble models. Actually, ensembles generated by TraDES and Flexible-Meccano are usually filtered and refined based on experimental data using computational tools such as ENSEMBLE [85], ASTEROIDS [86], EOM [87, 88] or the Maximum Occurrence [89, 90]. Integrative approaches, combining several complementary experimental and computational methods, are applied to derive more accurate structural models of IDPs/IDRs and their complexes (e.g., [91, 92]).

3. Interactions of structured domains mediated or regulated by disordered linkers

The majority of proteins in prokaryotes and eukaryotes are composed of several domains connected by linkers [93]. Domain-linker-domain (DLD), illustrated in the Figure 1.A, is the most common architecture, but more complex combinations of globular domains connected by flexible linkers exist. Although linkers can be very long, they typically involve from 2 up to ~ 30 residues [94, 95],

displaying high levels of flexibility and absence of permanent secondary structure. MD simulations in combination with ^{15}N NMR relaxation experiments have shown that this flexibility occurs in a broad range of timescales [55].

Linkers are not mere connectors between domains. Indeed, their length and sequence have been evolutionarily tailored to play key functional roles, being frequently involved in allosteric mechanisms [96, 97, 98]. One of the main advantages of this architecture is their capacity to enhance the effective local concentration, C_{eff} , of the linked domains, thus promoting intra- or intermolecular interactions (Figure 1.A₁-A₂). They are also key components in signaling processes: linkers can propagate conformational changes in one domain, e.g. induced by ligand binding, to the other domain, which may activate or inhibit other interactions (Figure 1.A₃). Below, we present examples of functional roles of linkers, and discuss how they have been investigated using various computational approaches.

3.1. Linkers in bi-specific antibodies

The role of linkers to enhance C_{eff} , as well as their effects on stability, affinity and activity, have been of particular interest in the context of bi-specific antibodies conceived from the combination of different antibodies or antibody fragments [99]. These engineered molecules have a great potential for diagnostic and therapeutic applications. The simplest and most common architecture, called single-chain variable domain (scFv) format, consists of antigen-binding sites of two antibodies connected through a linker. Theoretical methods based on simple worm-like models have been proposed to investigate the binding affinity of these systems [100], allowing to establish a relationship between the linker length and C_{eff} . However, predictions provided by such simple models can be inaccurate since they do not consider sequence-dependent structural properties of the linker and disregard possible interactions with the domains. Both, linker sequence and interactions, have been shown to be important for the conformational preferences of multi-domain proteins [101, 102]. Therefore, more detailed models are required for their investigation. In their study, Mittal *et al.* [102]

performed simulations using the ABSINTH together with an MC-based method called Hamiltonian Switch Metropolis Monte Carlo (HS-MMC) [103] specially developed to enhance sampling of IDRs connected to a folded domain. Although only relatively small artificial constructs involving SH3 and WW domains were used in this study, the approach and the conclusions can be generalized to other systems, including scFvs. For this type of systems, perturbation-response methods are a valuable tool to investigate the dynamical coupling between the two complementarity-determining regions (CDR), as well as the role of the linker in this mechanism. As an interesting example of such methods, Ettayapuram-Ramaprasad *et al.* [104] proposed an implementation based on an effective Hessian matrix computed from all-atom MD simulations. This Hessian matrix represents an ensemble-based elastic network that captures collective motions, from which the effect of local perturbations can be exhaustively investigated.

3.2. Linkers in multi-domain enzymes

Multi-domain enzymes are another type of proteins for which the study of the functional roles of linkers has attracted interest over the past two decades. MD simulations have been widely used for this purpose. For instance, standard all-atom MD protocols with simulation times of 20 ns were used to investigate the role of the linker in cullin-RING E3 ubiquitin ligases [105], unveiling that allosterically controlled linker motions modulate the distance between the domains, and therefore the ubiquitin transfer reactions. Nevertheless, these types of “basic” techniques cannot be applied to investigate thermodynamic and kinetic properties that would require extremely long simulations. CG models and enhanced sampling methods are the natural alternatives in this case. As an example, Li *et al.* [106] assessed the essential role of disordered linkers in allosteric regulation processes using a Gō-like model and umbrella sampling combined with a theoretical thermodynamic analysis. Their results suggested that the influence of the linker can be characterized by a C_{eff} that depends on the linker length and flexibility.

3.3. *The case of bimodular cellulases*

Numerous studies of multi-domain proteins involving flexible linkers are based on a combination of experimental and computational methods. Bimodular cellulases composed of covalently bound catalytic and cellulose-binding modules can be considered as a typical example. For instance, structural properties of a long disordered linker, containing 88 residues, in an artificial protein conceived from two natural cellulases were investigated by SAXS combined with molecular modeling tools [107]. More precisely, high-temperature MD simulations were applied as a conformational sampling technique, and a subset of the resulting models was selected to collectively fit the experimental data. Results of this study showed that the linker does not behave like a pure random coil, and suggest that the structural properties of the linker are essential for the function of these bimodular enzymes. Similar results have been observed in other studies combining SAXS and theoretical approaches [108, 109]. Moreover, bioinformatics analyses showed that sequence features are conserved in different families of bimodular cellulase enzymes, and suggest that the linker length has been evolutionarily optimized based on the type of the connected domains [110]. In this study, the authors also applied all-atom replica-exchange MD simulations together with circular dichroism to investigate the effects of glycosylation in the linker. Results of their analysis showed that the linkers are not rigidified by the addition of mono- or disaccharides, although they tend to adopt more extended conformations. Overall, this work demonstrated that linker length and composition is important for the activity of these enzymes, but a more clear description of functional roles remained to be elucidated. One of these roles was revealed by μ s-scale all-atom MD simulations, showing that glycosylated linkers bind dynamically and non-specifically to the cellulose surface [111]. The predicted enhancement of binding affinity due to the linker was confirmed experimentally. The importance of the linker for the processivity in cellulases, as well as in other DLD enzymes, has been investigated using bioinformatics tools and a statistical kinetic model [112]. Results of this theoretical work suggested that processivity may result from the kinetic bias of binding due to spatial constraints imposed

by the linker, which favors rebinding over full release of the substrate. They also show that the linker length and flexibility have been finely tuned through evolution to optimize this process.

4. Interactions between disordered regions and structured domains

The interaction between IDPs/IDRs and their globular partners is very often mediated by SLiMs inserted into disordered chains [15] (see Section 1 for additional details about SLiMs). In the unbound form, SLiMs can be pre-structured, reducing the entropic cost of the interaction and, as a consequence, tuning its thermodynamics [113, 114]. The inherent flexibility enables a single SLiM to recognize multiple partners with different structures and affinities (Fig. 2), with p53 being the most notorious example of this promiscuity [4]. Several proteins contain successive SLiMs and can be perceived as molecular platforms that bring to proximity different proteins involved in the same metabolic or signaling pathway to form high-order molecular assemblies [115]. For instance, this capacity is exploited by nuclear receptor co-regulators to assemble a large number of proteins to trigger gene transcription (see below), or by viruses to hijack the eukaryotic translational machinery [116]. In this section, we will describe how computational methods have helped to understand SLiM recognition events. Then, we describe the architectures emerging when several adjacent SLiMs recognize one or multiple sites in the globular partner.

4.1. Modeling partner recognition by short linear motifs

MD simulations have emerged as a powerful tool to study binding modes of IDPs. MD simulations are especially well-suited when the recognition and binding to the partner is achieved by SLiMs since in these cases the computational effort can be reduced by simulating only a small fragment of the IDP. In many the cases, high-resolution structures of the bound form are available from X-ray crystallography or NMR. Alternatively, experimentally-assisted computational docking with programs such as FlexPepDock [45], HADDOCK [46, 117] or ID-PLZerD [47] can be used to model the SLiM in the bound form. MD studies of

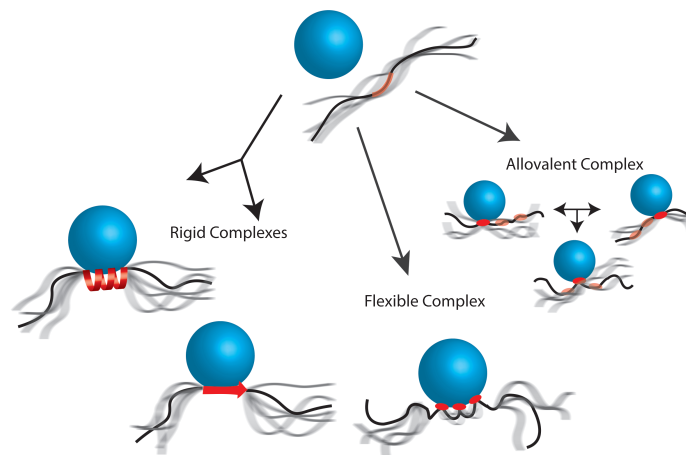


Figure 2: Representation of the different scenarios when an IDP interacts with a globular domain. Upon binding the interacting SLiM can adopt a rigid structure in a folding-upon-binding process. The final structure of this rigid segment, depending on the properties of the receptor, can be a canonical secondary structure (α -helix or β -strand), or adopt a coil conformation. Note that the interaction region of the IDP can be pre-structured in the unbound form, tuning the thermodynamic stability of the complex. The complex can also be dynamic, displaying multiple weak specific interactions that bind and unbind continuously while maintaining the overall architecture of the complex. Allovalent complexes occur when several SLiMs adjacently positioned in the chain can interact with a single receptor site and the bound conformation is continuously exchanging.

partner recognition by IDPs have primarily centered on discriminating between two mechanistically different binding modes: *conformational selection*, when the preformed bound conformation is a requirement for binding, and *induced fit*, when the optimal conformation is only adopted upon binding. For instance, the structural ensembles of Gab2 in the unbound state as well as in complex with Grb2 were generated using MD simulation with NMR-derived backbone chemical shifts as restraints [118]. Interestingly, it was observed that the secondary structure elements involved in recognition and binding of the partner were already present in the unbound state as well, albeit transiently. Disruption of these secondary structure elements resulted in an affinity reduction, establishing Gab2-Grb2 interaction as a typical example of *conformational se-*

lection. On the other hand, umbrella-sampling all-atom and coarse-grained MD simulations to study the binding between c-myb and KIX revealed a different scenario [119]. It was observed that the probability of crossing the transition state and the time required to do so did not depend on the structuration of c-myb at the beginning of the simulations, indicating that both unstructured and structured c-myb were capable of binding KIX with comparable rates. It was also noted that the transition state ensemble was heterogeneous with a wide diversity of c-myb conformations. A yet different mode of binding was observed for p53-MDM2 binding using very long unbiased MD simulations and Markov State Models (MSMs) [120]. In this case, binding almost always preceded folding, providing a classic example of ‘fly-casting’ followed by induced fit.

Another example of simultaneous binding and folding was described by Robustelli *et al.* for the interaction between the α -helical molecular recognition element (α -MoRE) of the intrinsically disordered C-terminal domain of the measles virus nucleoprotein (NTAIL) and the X domain (XD) of the phosphoprotein of the same virus using unbiased MD simulations [121]. As in the case of c-myb-KIX complex, the transition state was found to be highly heterogeneous. An interesting observation, however, was that if the α -MoRE formed long helices in the beginning of the binding event, it actually unfolded before forming the additional intermolecular contacts of the native conformation. This is in contrast to the conformational selection phenomenon observed for Gab2-Grb2. It was concluded that there was no clear temporal separation between binding and folding events as observed in other cases [122].

The above-described interactions can also occur intramolecularly if a disordered tail recognizes the globular domain to which it is attached (see Figure 1.B₁) [123, 124, 125]. Due to the concomitant increase of the C_{eff} , this architecture enables interactions that would have a very low affinity in an intermolecular scenario. Furthermore, the inherent flexibility of the resulting loop-like fragment between the the domain and the binding motif decreases the entropic cost of the interaction [126]. An example of such intramolecular

interaction is the auto-inhibition of DNA binding activity of Ets1 by its disordered C-terminal IDR having a Serine rich region (SRR) that can be phosphorylated [127]. Kasahara *et al.* used high-temperature canonical MD simulations to generate a wide range of structures which were then used to seed multi-canonical MD simulations for enhanced sampling. The simulations showed an increased number of contacts between the phosphorylated SRR and a helix in the core of the protein which is responsible for DNA binding, compared to non-phosphorylated SRR indicating a direct competitive mode of inhibition. Free energy surface analyses based on Principal Component Analysis (PCA) followed by clustering of conformations showed that these auto-inhibitory states existed in the non-phosphorylated state as well but their population was significantly increased upon phosphorylation due to alteration of the free energy landscape of Ets1.

4.2. Modeling allovalent complexes

Allovalent interactions occur when multiple similar (or equivalent) SLiMs are adjacently found in the same protein and interact with a partner with a single interaction site [19, 18] (see Figure 1.C₂). These polyvalent proteins enable a special type of fuzzy complex in which the different SLiMs alternatively recognize the partner and dynamically exchange their position from unbound to bound forms (Fig. 2). This competition of weak interactions for the same binding site increases the overall stability of the complex through cooperative effects that cannot be accounted for by traditional thermodynamic models [128, 129]. The continuous binding-dissociation-rebinding processes are very difficult to model, hampering the deep understanding of the structural and kinetic signatures of allovalency.

The interaction of phosphorylated Sic1 (pSic1) with *cdc4* is the prototypical example of allovalent complex. Sic1 contains nine similar CDK phosphorylation sites spread along the chain that can interact with the Cdc4 [130]. Interestingly, the increase of the affinity is not linear with the number phosphorylated sites, and the K_d reaches the submicromolar range only in the presence of at least 6 of

them [131]. This non-linear cooperative mechanism makes Sic1 extremely sensitive to the cellular level of the Cdk kinase [132]. Structural ensembles of Sic1 and pSic1 have been determined by combining NMR and SAXS data, which were integrated using the program ENSEMBLE [133]. A simplistic model of the allovalent complex was built by docking the ensemble of the unbound pSic1 to Cdc4 using the site-specific fraction of bound form determined by NMR and the crystallographic structure of Cdc4 with a model peptide. Although this model provides some insights into the binding mode, the thermodynamic and kinetic features of the complex remain elusive, requiring more advanced computational tools. MD simulations were performed to understand the allovalent recognition of a fragment of the nuclear pore complex (NPC) protein Nup135 and importin- β [134]. Like many other NPC proteins, Nup135 contains multiple FG dipeptides inserted in the sequence that, by weakly interacting with specific proteins, facilitate their translocation to the nucleus. Individual conformations of Nup135 derived from unbiased MD simulations were collected, mixed with importin- β and submitted to a 2 μ s MD simulation. The specific association of the two proteins was repeatedly observed along the trajectory, with the FG-repeats docking into previously identified binding pockets on the surface of importin- β [135]. Although the structural details of the FG recognition could be observed, the limited sampling hampered the extraction of the site-exchange kinetics and the evaluation of the differences between the alternative sites.

4.3. Modeling partner recognition by different short linear motifs

A different scenario occurs when multiple adjacent SLiMs can recognize the same globular partner through different anchoring points. In these circumstances, the disordered chain forms a long flexible loop-like structure that connects the bound segments (see Figure 1.C₃). This recognition mechanism is often associated to cooperative binding through the increase of the C_{eff} of other SLiM(s) when one or more SLiM(s) are already bound. Note that this mechanism is similar to the case of disordered linkers connecting globular domains explained in Section 3. Organisms have developed these complex regulation

mechanisms in order to modulate biological outputs. There are many examples of interactions that involve multisites, but very few of them provide a structural characterization of such complexes. Thus, the complexity of the whole system, including the interplay of the different interacting regions, often remains undescribed.

Complexes involving disordered co-regulators and homo- or heterodimeric nuclear receptor (NR) that regulate gene transcription are prototypical examples of the C_3 scenario. The interaction motifs of co-activators and co-repressors, called NR-boxes, share LxxLL and LxxI/HIxxI/L consensus sequence, respectively. Intriguingly, co-regulators contain a different number of consecutive NR-boxes depending on the organism and can potentially recognize the two binding sites of the NR dimers. For this multisite binding, the balance between an asymmetric model, where a single NR anchoring point is occupied, and a deck model, where both anchoring points are engaged, will depend on the affinity of the individual NR-boxes and the effective concentration dictated by the number and distance between the SLiMs. For the specific case of NRs, local affinities are modulated by endogenous ligands. The complex between the co-repressor N-CoR_{NID} with the RXR/RAR NR heterodimer could not be fully characterized at the residue level by NMR due to chemical exchange observed in the interacting regions [136]. In order to have a global picture of the complex, all-atom models of N-CoR_{NID} were generated using Flexible-Meccano [73, 74] and docked to one site of the heterodimer using the crystallographic structure as a template, representing the asymmetric model. To represent the deck model, with the two NR-boxes simultaneously bound to the heterodimer, steered MD simulations were performed on some conformations of the asymmetric ensemble forcing the second NR site to dock on the other face of the NR. By comparing the averaged SAXS profiles computed from both ensembles with the experimental one, the relative populations of the two binding modes in the apo form and in the presence of NR ligands were determined [136]. For the case of co-activators, no detailed model of the complexes has been proposed, although the presence of simultaneous binding has been demonstrated [137, 138, 139]. Inter-

estingly, for TIF2_{NRID} co-activator, NMR experiments highlighted the involvement of TIF2_{NRID} NR-box2 flanking region in its interaction with RXR/RAR heterodimer. The specific fragment encompassing NR-box2 and its flanking ordered region was co-crystallized with RAR bound to an agonist, and revealed an interacting helix turn helix motif of the TIF2_{NRID} fragment on the RAR surface [139]. The exact role of this flanking region in the recognition mechanism and the effects on the overall arrangement of the complex remain to be deciphered. Again, computational approaches should play a pivotal role to address these questions.

Another example concerns the interaction of a 60-residue long fragment of the tumor-suppressor p53¹⁻⁶⁰ with the metastasis-associated S100A4 protein through three anchoring points [140]. This study combined NMR data with MD simulations to determine the structure and dynamics of this fuzzy complex. The fact that the linkers between the three interaction motifs are short makes the modeling of the system less complicated. Indeed, the conformational sampling of long flexible loops connecting simultaneously bound SLiMs is one of the remaining challenges in the field. Although numerous methods have been reported for loop modeling in folded proteins [141, 97, 142], existing approaches mainly aim at predicting the most likely loop conformation(s) rather than exhaustively sampling the conformational space of the loop. Moreover, only a few of these methods remain computationally efficient when the loop length exceeds 15 residues. One of them is a robotics-inspired method that exploits a large structural database of three-residue fragments [143]. First tests with this method applied to IDPs show its ability to rapidly generate conformational ensemble models of loops involving around 100 residues (unpublished work).

5. Extreme fuzzy complexes and phase separation behavior

Several IDPs can also interact with each other. The association can give rise to highly disordered complexes [144] (illustrated in Figure 1.D₂ and D₃) or to rigid particles, such as amyloids (Figure 1.D₁). In this last case, large aggregates

are formed by the perfect arrangement of chains stabilized by a dense network of hydrogen bonds. This case will not be described here, and the reader is referred to other publications [145, 146, 147]. At the other extreme of flexibility, recent publications describe the formation of high-affinity complexes between two IDPs that retain their flexibility upon binding [20, 148]. For the case of Borgia *et al.* [20], this new kind of biomolecular interaction can be explained by the large opposite electrostatic charges of the two proteins, histone H1 and its nuclear chaperone prothymosin- α . The integration of NMR and smFRET data into one-bead-per-residue CG simulations unveiled that the complex was maintained by multiple long-range electrostatic interactions without the need for defined binding sites and specific interactions. Interestingly, ternary complexes displaying a high exchange rate are formed at high concentrations [149]. The lack of specificity in the interactions causes this phenomenon and triggers the formation of large oligomers, a phenomenon that is reminiscent of liquid-liquid phase separation (LLPS).

Multiple pieces of evidence indicate that dynamical, multivalent interactions between IDRs/IDPs are major drivers of cellular LLPS processes and provide the structural scaffold for the so-called membrane-less organelles [22, 150]. Remarkably, the structural and functional characterization of these condensates is attracting ever-growing attention since they are currently recognized to play a major role in organizing cellular biochemistry [21, 151]. Computational approaches have the potential to play a key role in this challenge, given the difficulties in tackling the daunting complexity of these biomolecular assemblies with standard structural biology techniques and/or polymer physics theories [150, 38]. In particular, molecular simulations can provide access to elusive structural details of the condensates and complement theoretical and experimental investigations of the molecular grammar governing LLPS [152, 153, 154] with the final aim of establishing sequence-structure-function relations.

Not surprisingly, the length and time scales associated with cellular LLPS, which are collective processes involving intermolecular interactions among a large number of large-sized biomolecules, have favored the development and ap-

plications of suitable CG molecular models. In this respect, CG models based on one-bead-per-residue description have been shown to provide a reasonable compromise of accuracy and computational efficiency and are a popular choice for simulating the LLPS equilibria of flexible proteins [37, 38, 39, 40, 41]. Most accurate versions of these models explicitly take into account the protein sequence and rely on inter-residue energy functions that include implicit-solvent Debye-Hückel electrostatics and contact potentials accounting for excluded volume and short-range attraction. The latter terms are defined according to hydrophobicity scales or statistical potentials and tuned to reproduce experimental structural data or affinities [155, 156]. While the quantitative predictive capabilities of one-bead-per-residue potential should not be overstated [157], this approach has been successfully applied to shed light on how the protein sequence determines the phase behavior of IDPs, as well as the structural and dynamical properties of condensed phase [156, 154, 158]. Furthermore, CG simulations at this level of resolution can be easily extended to include folded domains [159], post-translational modifications [160], thermoresponsive behavior [161] and interactions with RNA molecules [162]. Moving to higher-resolution models, a recent study indicated that the popular MARTINI CG force-field, which relies on a four-atoms to one-bead mapping and an explicit solvation model, can accurately describe the condensation of FUS prion-like domain, upon a fine tuning of its energy function against experimental transfer free-energies [163]. Conversely, ultra-coarse grained simulations, where a single bead may represent a protein domain or an entire biomolecule, have been successfully applied to get some insight into the internal organization of multi-component mixtures that mimic more closely the complexity of cellular condensates [164, 165, 166, 167].

So far, the role of atomistic MD in this field has been rather limited due to the demanding computational requirements of this approach, which make the direct simulation of phase separation processes unfeasible with present-day computational resources. Nevertheless, recent studies have indicated novel strategies to take advantage of all-atom, explicit-solvent MD simulations based on accurate last-generation force fields in the characterization of biomolecular LLPS. No-

tably, MD simulations of protein fragments at high-concentration were used to dissect the molecular interactions driving the LLPS with a “divide-and-conquer” strategy and they provided results in good agreement with NMR and mutagenesis data with a limited computational cost [168]. Furthermore, a high-resolution picture of protein dynamics in the condensed phase was obtained by generating an initial CG configuration of phase-separated proteins, which was then mapped back to all-atom resolution and simulated in the microsecond timescale thanks to a specialized supercomputer [169].

6. Concluding Remarks

The biological relevance of IDPs/IDRs underlines the importance of having detailed structural models of this class of proteins and their complexes. These models guarantee a molecular perspective of key cellular processes and eventual rational interventions with pharmacological aims [170]. The co-existence of an astronomical number of conformations and the averaged nature of the experimental data that can be recorded for IDPs make the use of computational methods unavoidable. The immense challenges in the field are exemplified in the study of liquid-like droplets, which have attracted the interest of a large community from diverse scientific domains. These highly concentrated protein condensates are inherently disordered and display multivalent, weak intermolecular interactions that are modulated by external parameters such as pH, temperature or phosphorylation states. Therefore, they present multiple challenges for computational modeling.

The growing interest of the structural bioinformatics community to overcome challenges posed by IDPs/IDRs is encouraging. The improvement in the force fields, for both all-atom and CG simulations, to adapt them to disordered states, the development of enhanced sampling strategies, as well as the generalization of parallelized software and the use of GPUs are the most prominent hints of these developments. The increase in the number of experimental studies focusing on IDPs/IDRs is also crucial as they continue identifying novel

biological mechanisms. Moreover, databases and repositories assembling experimental and omics data improve our structural and functional knowledge of these proteins, and provide new opportunities to develop and validate the theoretical methods [171, 42]. This new data is rich in information and can be used, for instance, to improve current force fields, or can be exploited to conceive more accurate conformational sampling methods [75]. The use of data mining and machine learning methods to analyze and exploit relevant information from these databases is a very promising avenue for the improvement of predictive molecular modeling approaches and for the development of new tools to tackle the challenging questions posed by disordered proteins and their complexes.

CRedit authorship contribution statement

Ilinka Clerc: Writing – Original Draft Preparation, Visualization. **Amin Sagar:** Writing – Original Draft Preparation, Writing - Review & Editing, Visualization. **Alessandro Barducci:** Writing – Original Draft Preparation, Writing - Review & Editing. **Nathalie Sibille:** Writing – Original Draft Preparation. **Pau Bernadó:** Supervision, Writing – Original Draft Preparation, Writing - Review & Editing. **Juan Cortés:** Supervision, Writing – Original Draft Preparation, Writing - Review & Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been supported by the French National Research Agency (ANR) through grants ANR-17-CE11-0022-01, ANR-19-PI3A-0004, and Labex EpiGenMed, an “*Investissements d’avenir*” program (ANR-10-LABX-12-01).

The *Centre de Biochimie Structurale* (CBS) is a member of France-BioImaging (FBI) and the French Infrastructure for Integrated Structural Biology (FRISBI), 2 national infrastructures supported by the ANR (ANR-10-INBS-04-01 and ANR-10-INBS-05, respectively).

References

- [1] H. J. Dyson, P. E. Wright, Intrinsically unstructured proteins and their functions, *Nat Rev Mol Cell Biol* 6 (2005) 197–208. doi:10.1038/nrm1589.
- [2] C. J. Oldfield, A. K. Dunker, Intrinsically disordered proteins and intrinsically disordered protein regions, *Annu Rev Biochem* 83 (1) (2014) 553–584. doi:10.1146/annurev-biochem-072711-164947.
- [3] H. Xie, S. Vucetic, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, V. N. Uversky, Z. Obradovic, Functional anthology of intrinsic disorder. 1. biological processes and functions of proteins with long disordered regions, *J Proteome Res* 6 (5) (2007) 1882–1898. doi:10.1021/pr060392u.
- [4] R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, M. M. Babu, Classification of intrinsically disordered regions and proteins, *Chem Rev* 114 (13) (2014) 6589–6631. doi:10.1021/cr400525m.
- [5] P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation, *Nat Rev Mol Cell Biol* 16 (2015) 18–29. doi:10.1038/nrm3920.
- [6] V. Csizmok, A. V. Follis, R. W. Kriwacki, J. D. Forman-Kay, Dynamic protein interaction networks and new structural paradigms in signaling, *Chem Rev* 116 (11) (2016) 6424–6462. doi:10.1021/acs.chemrev.5b00548.

- [7] P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, A. K. Dunker, Sequence complexity of disordered protein, *Proteins* 42 (1) (2001) 38–48. doi:10.1002/1097-0134(20010101)42:1<38::AID-PROT50>3.0.CO;2-3.
- [8] V. N. Uversky, Natively unfolded proteins: a point where biology waits for physics, *Protein Sci* 11 (2002) 739–756. doi:10.1110/ps.4210102.
- [9] P. Mier, L. Paladin, S. Tamana, S. Petrosian, B. Hajdu-Soltész, A. Urbanek, A. Gruca, D. Plewczynski, M. Grynberg, P. Bernadó, Z. Gáspári, C. A. Ouzounis, V. J. Promponas, A. V. Kajava, J. M. Hancock, S. C. E. Tosatto, Z. Dosztanyi, M. A. Andrade-Navarro, Disentangling the complexity of low complexity proteins, *Brief Bioinform* 21 (2) (2019) 458–472. doi:10.1093/bib/bbz007.
- [10] S. Vucetic, C. J. Brown, A. K. Dunker, Z. Obradovic, Flavors of protein disorder, *Proteins* 52 (4) (2003) 573–584. doi:10.1002/prot.10437.
- [11] Y. Liu, X. Wang, B. Liu, A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction, *Brief Bioinform* 20 (1) (2017) 330–346. doi:10.1093/bib/bbx126.
- [12] J. T. Nielsen, F. A. A. Mulder, Quality and bias of protein disorder predictors, *Sci Rep* 9 (2019) 5137. doi:10.1038/s41598-019-41644-w.
- [13] P. Tompa, The interplay between structure and function in intrinsically unstructured proteins, *FEBS Lett* 579 (15) (2005) 3346–3354. doi:10.1016/j.febslet.2005.03.072.
- [14] P. Tompa, E. Schad, A. Tantos, L. Kalmar, Intrinsically disordered proteins: emerging interaction specialists, *Curr Opin Struct Biol* 35 (2015) 49 – 59. doi:10.1016/j.sbi.2015.08.009.

- [15] N. E. Davey, K. Van Roey, R. J. Weatheritt, G. Toedt, B. Uyar, B. Altenberg, A. Budd, F. Diella, H. Dinkel, T. J. Gibson, Attributes of short linear motifs, *Mol BioSyst* 8 (2012) 268–281. doi:10.1039/C1MB05231D.
- [16] K. Van Roey, B. Uyar, R. J. Weatheritt, H. Dinkel, M. Seiler, A. Budd, T. J. Gibson, N. E. Davey, Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation, *Chem Rev* 114 (13) (2014) 6733–6778. doi:10.1021/cr400585q.
- [17] M. Kumar, M. Gouw, S. Michael, H. Sámano-Sánchez, R. Pancsa, J. Glavina, A. Diakogianni, J. A. Valverde, D. Bukirova, J. Čalyševa, N. Palopoli, N. E. Davey, L. B. Chemes, T. J. Gibson, ELM—the eukaryotic linear motif resource in 2020, *Nucleic Acids Res* 48 (D1) (2019) D296–D306. doi:10.1093/nar/gkz1030.
- [18] M. Fuxreiter, Classifying the binding modes of disordered proteins, *Int J Mol Sci* 21 (22) (2020). doi:10.3390/ijms21228615.
- [19] J. G. Olsen, K. Teilum, B. B. Kragelund, Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness, *Cell Mol Life Sci* 74 (2017) 3175–3183. doi:10.1007/s00018-017-2560-7.
- [20] A. Borgia, M. Borgia, K. Bugge, V. Kissling, P. Heidarsson, C. Fernandes, A. Sottini, A. Soranno, K. Buholzer, D. Nettels, B. B. Kragelund, R. Best, B. Schuler, Extreme disorder in an ultrahigh-affinity protein complex, *Nature* 555 (2018) 61–66. doi:10.1038/nature25762.
- [21] Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease, *Science* 357 (6357) (2017). doi:10.1126/science.aaf4382.
- [22] L. Bergeron-Sandoval, N. Safaee, S. Michnick, Mechanisms and consequences of macromolecular phase separation, *Cell* 165 (2016) 1067–1079. doi:10.1016/j.cell.2016.05.026PANIST.

- [23] M. Miskei, A. Horvath, M. Vendruscolo, M. Fuxreiter, Sequence-based prediction of fuzzy protein interactions, *J Mol Biol* 432 (7) (2020) 2289 – 2303. doi:10.1016/j.jmb.2020.02.017.
- [24] P. Tompa, N. Davey, T. Gibson, M. Babu, A million peptide motifs for the molecular biologist, *Mol Cell* 55 (2) (2014) 161–169. doi:10.1016/j.molcel.2014.05.032.
- [25] V. Csizmok, J. D. Forman-Kay, Complex regulatory mechanisms mediated by the interplay of multiple post-translational modifications, *Curr Opin Struct Biol* 48 (2018) 58 – 67. doi:10.1016/j.sbi.2017.10.013.
- [26] A. Bah, J. D. Forman-Kay, Modulation of intrinsically disordered protein function by post-translational modifications, *J Biol Chem* 291 (2016) 6696–6705. doi:10.1074/jbc.R115.695056.
- [27] V. N. Uversky, C. J. Oldfield, A. K. Dunker, Intrinsically disordered proteins in human diseases: Introducing the d2 concept, *Annu Rev Biophys* 37 (1) (2008) 215–246. doi:10.1146/annurev.biophys.37.032807.125924.
- [28] Y. Pérez, M. Maffei, A. Igea, I. Amata, M. Gairí, A. R. Nebreda, P. Bernadó, M. Pons, Lipid binding by the unique and SH3 domains of c-Src suggests a new regulatory mechanism, *Sci Rep* 3 (2013) 1295. doi:10.1038/srep01295.
- [29] S. Milles, N. Salvi, M. Blackledge, M. R. Jensen, Characterization of intrinsically disordered proteins and their dynamic complexes: From in vitro to cell-like environments, *Prog Nucl Magn Reson Spectrosc* 109 (2018) 79 – 100. doi:10.1016/j.pnmrs.2018.07.001.
- [30] H. J. Dyson, P. E. Wright, Nmr illuminates intrinsic disorder, *Curr Opin Struct Biol* 70 (2021) 44–52. doi:https://doi.org/10.1016/j.sbi.2021.03.015.

- [31] T. N. Cordeiro, F. Herranz-Trillo, A. Urbanek, A. Estaña, J. Cortés, N. Sibille, P. Bernadó, Small-angle scattering studies of intrinsically disordered proteins and their complexes, *Curr Opin Struct Biol* 42 (2017) 15 – 23. doi:10.1016/j.sbi.2016.10.011.
- [32] V. Receveur-Brechot, D. Durand, How random are intrinsically disordered proteins? a small angle scattering perspective, *Curr Protein Pept Sci* 13 (1) (2012) 55–75. doi:10.2174/138920312799277901.
- [33] E. D. Holmstrom, A. Holla, W. Zheng, D. Nettels, R. B. Best, B. Schuler, Chapter ten - accurate transfer efficiencies, distance distributions, and ensembles of unfolded and intrinsically disordered proteins from single-molecule fret, in: E. Rhoades (Ed.), *Intrinsically Disordered Proteins*, Vol. 611 of *Methods in Enzymology*, Academic Press, 2018, pp. 287 – 325. doi:10.1016/bs.mie.2018.09.030.
- [34] E. Ravera, L. Sgheri, G. Parigi, C. Luchinat, A critical assessment of methods to recover information from averaged data, *Phys Chem Chem Phys* 18 (2016) 5686–5701. doi:10.1039/C5CP04077A.
- [35] P. Bernadó, M. Blackledge, Proteins in dynamic equilibrium, *Nature* 468 (2010) 1046–1048. doi:10.1038/4681046a.
- [36] T. N. Cordeiro, P.-C. Chen, A. De Biasio, N. Sibille, F. J. Blanco, J. S. Hub, R. Crehuet, P. Bernadó, Disentangling polydispersity in the PCNAp15PAF complex, a disordered, transient and multivalent macromolecular assembly, *Nucleic Acids Res* 45 (3) (2016) 1501–1515. doi:10.1093/nar/gkw1183.
- [37] R. B. Best, Computational and theoretical advances in studies of intrinsically disordered proteins, *Curr Opin Struct Biol* 42 (2017) 147 – 154. doi:10.1016/j.sbi.2017.01.006.
- [38] K. M. Ruff, R. V. Pappu, A. S. Holehouse, Conformational preferences and phase behavior of intrinsically disordered low complexity sequences:

- insights from multiscale simulations, *Curr Opin Struct Biol* 56 (2019) 1 – 10. doi:10.1016/j.sbi.2018.10.003.
- [39] K. Kasahara, H. Terazawa, T. Takahashi, J. Higo, Studies on molecular dynamics of intrinsically disordered proteins and their fuzzy complexes: A mini-review, *Comput Struct Biotechnol J* 17 (2019) 712 – 720. doi:10.1016/j.csbj.2019.06.009.
- [40] S. Bhattacharya, X. Lin, Recent advances in computational protocols addressing intrinsically disordered proteins, *Biomolecules* 9 (4) (2019). doi:10.3390/biom9040146.
- [41] J.-E. Shea, R. B. Best, J. Mittal, Physics-based computational and theoretical approaches to intrinsically disordered proteins, *Curr Opin Struct Biol* 67 (2021) 219–225. doi:10.1016/j.sbi.2020.12.012.
- [42] T. Lazar, E. Martínez-Pérez, F. Quaglia, A. Hatos, L. Chemes, J. A. Iserte, N. A. Méndez, N. A. Garrone, T. Saldaño, J. Marchetti, A. Rueda, P. Bernadó, M. Blackledge, T. N. Cordeiro, E. Fagerberg, J. D. Forman-Kay, M. Fornasari, T. J. Gibson, G.-N. W. Gomes, C. Gradinaru, T. Head-Gordon, M. R. Jensen, E. Lemke, S. Longhi, C. Marino-Buslje, G. Minervini, T. Mittag, A. Monzon, R. V. Pappu, G. Parisi, S. Ricard-Blum, K. M. Ruff, E. Salladini, M. Skepö, D. Svergun, S. Vallet, M. Varadi, P. Tompa, S. C. E. Tosatto, D. Piovesan, PED in 2021: A major update of the protein ensemble database for intrinsically disordered proteins, *Nucleic Acids Res* 49 (D1) (2020) D404–D411. doi:10.1093/nar/gkaa1021.
- [43] Z. Dosztányi, B. Mészáros, I. Simon, ANCHOR: web server for predicting protein binding regions in disordered proteins, *Bioinformatics* 25 (20) (2009) 2745–2746. doi:10.1093/bioinformatics/btp518.
- [44] I. Krystkowiak, N. E. Davey, SLiMSearch: a framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions, *Nucleic Acids Res* 45 (W1) (2017) W464–W469. doi:10.1093/nar/gkx238.

- [45] B. Raveh, N. London, L. Zimmerman, O. Schueler-Furman, Rosetta FlexPepDock ab-initio: Simultaneous folding, docking and refinement of peptides onto their receptors, *PloS One* 6 (4) (2011) 1–10. doi:10.1371/journal.pone.0018934.
- [46] C. Geng, S. Narasimhan, J. P. G. L. M. Rodrigues, A. M. J. J. Bonvin, Information-Driven, Ensemble Flexible Peptide Docking Using HADDOCK, Springer New York, New York, NY, 2017, pp. 109–138. doi:10.1007/978-1-4939-6798-8_8.
- [47] L. X. Peterson, A. Roy, C. Christoffer, G. Terashi, D. Kihara, Modeling disordered protein interactions from biophysical principles, *PLoS Comput Biol* 13 (4) (2017) 1–28. doi:10.1371/journal.pcbi.1005485.
- [48] U. R. Shrestha, J. C. Smith, L. Petridis, Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations, *Commun Biol* 4 (2021) 243. doi:10.1038/s42003-021-01759-1.
- [49] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, A. D. MacKerell, CHARMM36m: an improved force field for folded and intrinsically disordered proteins, *Nat Methods* 14 (2017) 71–73. doi:10.1038/nmeth.4067.
- [50] D. Song, R. Luo, H.-F. Chen, The IDP-specific force field ff14IDPSFF improves the conformer sampling of intrinsically disordered proteins, *J Chem Inf Model* 57 (5) (2017) 1166–1178, PMID: 28448138. doi:10.1021/acs.jcim.7b00135.
- [51] P. Robustelli, S. Piana, D. E. Shaw, Developing a molecular dynamics force field for both folded and disordered protein states, *Proc Natl Acad Sci USA* 115 (21) (2018) E4758–E4766. doi:10.1073/pnas.1800690115.
- [52] S. Piana, P. Robustelli, D. Tan, S. Chen, D. E. Shaw, Development of a force field for the simulation of single-chain proteins and protein-protein

- complexes, *J Chem Theory Comput* 16 (4) (2020) 2494–2507. doi:10.1021/acs.jctc.9b00251.
- [53] S. Piana, A. G. Donchev, P. Robustelli, D. E. Shaw, Water dispersion interactions strongly influence simulated structural properties of disordered protein states, *J Phys Chem B* 119 (16) (2015) 5113–5123. doi:10.1021/jp508971m.
- [54] R. B. Best, W. Zheng, J. Mittal, Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association, *J Chem Theory Comput* 10 (11) (2014) 5113–5124. doi:10.1021/ct500569b.
- [55] S. I. Virtanen, A. M. Kiirikki, K. M. Mikula, H. Iwai, O. H. S. Ollila, Heterogeneous dynamics in partially disordered proteins, *Phys Chem Chem Phys* 22 (2020) 21185–21196. doi:10.1039/D0CP03473H.
- [56] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, *Chem Phys Lett* 314 (1) (1999) 141–151. doi:10.1016/S0009-2614(99)01123-9.
- [57] H. Fukunishi, O. Watanabe, S. Takada, On the hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction, *J Chem Phys* 116 (20) (2002) 9058–9067. doi:10.1063/1.1472510.
- [58] A. Laio, M. Parrinello, Escaping free-energy minima, *Proc Natl Acad Sci USA* 99 (20) (2002) 12562–12566. doi:10.1073/pnas.202427399.
- [59] A. Barducci, M. Bonomi, M. Parrinello, Metadynamics, *WIREs Comput Mol Sci* 1 (5) (2011) 826–843. doi:10.1002/wcms.31.
- [60] G. Bussi, F. L. Gervasio, A. Laio, M. Parrinello, Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics, *J Am Chem Soc* 128 (41) (2006) 13435–13441. doi:10.1021/ja062463w.

- [61] S. Piana, A. Laio, A bias-exchange approach to protein folding, *J Phys Chem B* 111 (17) (2007) 4553–4559. doi:10.1021/jp0678731.
- [62] C. Abrams, G. Bussi, Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration, *Entropy* 16 (1) (2014) 163–199. doi:10.3390/e16010163.
- [63] Y. I. Yang, Q. Shao, J. Zhang, L. Yang, Y. Q. Gao, Enhanced sampling in molecular dynamics, *J Chem Phys* 151 (7) (2019) 070902. doi:10.1063/1.5109531.
- [64] J. E. Stone, J. C. Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco, K. Schulten, Accelerating molecular modeling applications with graphics processors, *J Comput Chem* 28 (16) (2007) 2618–2640. doi:10.1002/jcc.20829.
- [65] B. W. Zhang, D. Jasnow, D. M. Zuckerman, Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin, *Proc Natl Acad Sci USA* 104 (46) (2007) 18043–18048. doi:10.1073/pnas.0706349104.
- [66] M. Knott, R. B. Best, Discriminating binding mechanisms of an intrinsically disordered protein via a multi-state coarse-grained model, *J Chem Phys* 140 (17) (2014) 175102. doi:10.1063/1.4873710.
- [67] H. Wu, P. G. Wolynes, G. A. Papoian, AWSEM-IDP: A coarse-grained force field for intrinsically disordered proteins, *J Phys Chem B* 122 (49) (2018) 11115–11125. doi:10.1021/acs.jpcc.8b05791.
- [68] T. Bereau, M. Deserno, Generic coarse-grained model for protein folding and aggregation, *J Chem Phys* 130 (23) (2009) 235106. doi:10.1063/1.3152842.
- [69] D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman, S. J. Marrink, Improved pa-

- rameters for the martini coarse-grained protein force field, *J Chem Theory Comput* 9 (1) (2013) 687–697. doi:10.1021/ct300646g.
- [70] F. Klein, E. E. Barrera, S. Pantano, Assessing SIRAH’s Capability to Simulate Intrinsically Disordered Proteins and Peptides, *J Chem Theory Comput* 17 (2) (2021) 599–604. doi:10.1021/acs.jctc.0c00948.
- [71] A. Vitalis, R. V. Pappu, ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions, *J Comput Chem* 30 (5) (2009) 673–699. doi:https://doi.org/10.1002/jcc.21005.
- [72] H. J. Feldman, C. W. Hogue, Probabilistic sampling of protein conformations: New hope for brute force?, *Proteins* 46 (1) (2002) 8–23. doi:10.1002/prot.1163.
- [73] P. Bernadó, L. Blanchard, P. Timmins, D. Marion, R. W. H. Ruigrok, M. Blackledge, A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering, *Proc Natl Acad Sci USA* 102 (47) (2005) 17002–17007. doi:10.1073/pnas.0506202102.
- [74] V. Ozenne, F. Bauer, L. Salmon, J.-r. Huang, M. R. Jensen, S. Segard, P. Bernadó, C. Charavay, M. Blackledge, Flexible-Meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables, *Bioinformatics* 28 (11) (2012) 1463–1470. doi:10.1093/bioinformatics/bts172.
- [75] A. Estaña, N. Sibille, E. Delaforge, M. Vaisset, J. Cortés, P. Bernadó, Realistic ensemble models of intrinsically disordered proteins using a structure-encoding coil database, *Structure* 27 (2) (2019) 381–391.e2. doi:10.1016/j.str.2018.10.016.
- [76] B. Roux, J. Weare, On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method, *J Chem Phys* 138 (8) (2013) 084107. doi:10.1063/1.4792208.

- [77] W. Boomsma, J. Ferkinghoff-Borg, K. Lindorff-Larsen, Combining experiments and simulations using the maximum entropy principle, *PLoS Comput Biol* 10 (2) (2014) 1–9. doi:10.1371/journal.pcbi.1003406.
- [78] G. Hummer, J. Köfinger, Bayesian ensemble refinement by replica simulations and reweighting, *J Chem Phys* 143 (24) (2015) 243150. doi:10.1063/1.4937786.
- [79] N. Salvi, A. Abyzov, M. Blackledge, Multi-timescale dynamics in intrinsically disordered proteins from NMR relaxation and molecular simulation, *J Phys Chem Lett* 7 (13) (2016) 2483–2489. doi:10.1021/acs.jpcllett.6b00885.
- [80] M. Bonomi, C. Camilloni, A. Cavalli, M. Vendruscolo, MetaInference: A bayesian inference method for heterogeneous systems, *Sci Adv* 2 (1) (2016). doi:10.1126/sciadv.1501177.
- [81] M. Bonomi, G. T. Heller, C. Camilloni, M. Vendruscolo, Principles of protein structural ensemble determination, *Curr Opin Struct Biol* 42 (2017) 106–116. doi:10.1016/j.sbi.2016.12.004.
- [82] R. Rangan, M. Bonomi, G. T. Heller, A. Cesari, G. Bussi, M. Vendruscolo, Determination of structural ensembles of proteins: Restraining vs reweighting, *J Chem Theory Comput* 14 (12) (2018) 6632–6641. doi:10.1021/acs.jctc.8b00738.
- [83] J. Köfinger, L. S. Stelzl, K. Reuter, C. Allande, K. Reichel, G. Hummer, Efficient ensemble refinement by reweighting, *J Chem Theory Comput* 15 (5) (2019) 3390–3401. doi:10.1021/acs.jctc.8b01231.
- [84] S. Bottaro, T. Bengtson, K. Lindorff-Larsen, Integrating Molecular Simulation and Experimental Data: A Bayesian/Maximum Entropy Reweighting Approach, Springer US, New York, NY, 2020, pp. 219–240. doi:10.1007/978-1-0716-0270-6_15.

- [85] M. Krzeminski, J. A. Marsh, C. Neale, W.-Y. Choy, J. D. Forman-Kay, Characterization of disordered proteins with ENSEMBLE, *Bioinformatics* 29 (3) (2012) 398–399. doi:10.1093/bioinformatics/bts701.
- [86] G. Nodet, L. Salmon, V. Ozenne, S. Meier, M. R. Jensen, M. Blackledge, Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from nmr residual dipolar couplings, *J Am Chem Soc* 131 (49) (2009) 17908–17918. doi:10.1021/ja9069024.
- [87] P. Bernadó, E. Mylonas, M. V. Petoukhov, M. Blackledge, D. I. Svergun, Structural characterization of flexible proteins using small-angle x-ray scattering, *J Am Chem Soc* 129 (17) (2007) 5656–5664, PMID: 17411046. doi:10.1021/ja069124n.
- [88] G. Tria, H. D. T. Mertens, M. Kachala, D. I. Svergun, Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering, *IUCrJ* 2 (2) (2015) 207–217. doi:10.1107/S205225251500202X.
- [89] I. Bertini, A. Giachetti, C. Luchinat, G. Parigi, M. V. Petoukhov, R. Pierattelli, E. Ravera, D. I. Svergun, Conformational space of flexible biological macromolecules from average data, *J Am Chem Soc* 132 (38) (2010) 13553–13558. doi:10.1021/ja1063923.
- [90] M. Nagulapalli, G. Parigi, J. Yuan, J. Gsponer, G. Deraos, V. V. Bamm, G. Harauz, J. Matsoukas, M. R. R. de Planque, I. P. Gerotheranassis, M. M. Babu, C. Luchinat, A. G. Tzakos, Recognition pliability is coupled to structural heterogeneity: A calmodulin intrinsically disordered binding region complex, *Structure* 20 (3) (2012) 522–533. doi:10.1016/j.str.2012.01.021.
- [91] A. Borgia, W. Zheng, K. Buholzer, M. B. Borgia, A. Schüler, H. Hofmann, A. Soranno, D. Nettels, K. Gast, A. Grishaev, R. B. Best, B. Schuler, Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods, *J Am Chem Soc* 138 (36) (2016) 11714–11726. doi:10.1021/jacs.6b05917.

- [92] G.-N. W. Gomes, M. Krzeminski, A. Namini, E. W. Martin, T. Mittag, T. Head-Gordon, J. D. Forman-Kay, C. C. Gradinaru, Conformational ensembles of an intrinsically disordered protein consistent with nmr, saxs, and single-molecule fret, *J Am Chem Soc* 142 (37) (2020) 15697–15710. doi:10.1021/jacs.0c02088.
- [93] G. Apic, W. Huber, S. A. Teichmann, Multi-domain protein families and domain pairs: comparison with known structures and a random model of domain recombination, *J Struct Funct Genomics* 4 (2-3) (2003) 67–78. doi:10.1023/A:1026113408773.
- [94] R. A. George, J. Heringa, An analysis of protein domain linkers: their classification and role in protein folding, *Protein Eng Des Sel* 15 (11) (2002) 871–879. doi:10.1093/protein/15.11.871.
- [95] V. P. Reddy Chichili, V. Kumar, J. Sivaraman, Linkers in the structural biology of protein-protein interactions, *Protein Sci* 22 (2) (2013) 153–167. doi:10.1002/pro.2206.
- [96] B. Ma, C.-J. Tsai, T. Haliloğlu, R. Nussinov, Dynamic allostery: Linkers are not merely flexible, *Structure* 19 (7) (2011) 907 – 917. doi:10.1016/j.str.2011.06.002.
- [97] E. Papaleo, G. Saladino, M. Lambrugh, K. Lindorff-Larsen, F. L. Gervasio, R. Nussinov, The role of protein loops and linkers in conformational dynamics and allostery, *Chem Rev* 116 (11) (2016) 6391–6423. doi:10.1021/acs.chemrev.5b00623.
- [98] Q. Huang, M. Li, L. Lai, Z. Liu, Allostery of multidomain proteins with disordered linkers, *Curr Opin Struct Biol* 62 (2020) 175 – 182. doi:10.1016/j.sbi.2020.01.017.
- [99] U. Brinkmann, R. E. Kontermann, The making of bispecific antibodies, *mAbs* 9 (2) (2017) 182–212. doi:10.1080/19420862.2016.1268307.

- [100] H.-X. Zhou, Quantitative account of the enhanced affinity of two linked scfvs specific for different epitopes on the same antigen, *J Mol Biol* 329 (1) (2003) 1–8. doi:10.1016/S0022-2836(03)00372-3.
- [101] J. S. Klein, S. Jiang, R. P. Galimidi, J. R. Keeffe, P. J. Bjorkman, Design and characterization of structured protein linkers with differing flexibilities, *Protein Eng Des Sel* 27 (10) (2014) 325–330. doi:10.1093/protein/gzu043.
- [102] A. Mittal, A. S. Holehouse, M. C. Cohan, R. V. Pappu, Sequence-to-conformation relationships of disordered regions tethered to folded domains of proteins, *J Mol Biol* 430 (16) (2018) 2403 – 2421. doi:10.1016/j.jmb.2018.05.012.
- [103] A. Mittal, N. Lyle, T. S. Harmon, R. V. Pappu, Hamiltonian switch metropolis monte carlo simulations for improved conformational sampling of intrinsically disordered regions tethered to ordered domains of proteins, *J Chem Theory Comput* 10 (8) (2014) 3550–3562. doi:10.1021/ct5002297.
- [104] A. S. Ettayapuram Ramaprasad, S. Uddin, J. Casas-Finet, D. J. Jacobs, Decomposing dynamical couplings in mutated scFv antibody fragments into stabilizing and destabilizing effects, *J Am Chem Soc* 139 (48) (2017) 17508–17517. doi:10.1021/jacs.7b09268.
- [105] J. Liu, R. Nussinov, Molecular dynamics reveal the essential role of linker motions in the function of cullin–RING E3 ligases, *J Mol Biol* 396 (5) (2010) 1508 – 1523. doi:0.1016/j.jmb.2010.01.022.
- [106] M. Li, H. Cao, L. Lai, Z. Liu, Disordered linkers in multidomain allosteric proteins: Entropic effect to favor the open state or enhanced local concentration to favor the closed state?, *Protein Sci* 27 (9) (2018) 1600–1610. doi:10.1002/pro.3475.

- [107] I. von Ossowski, J. T. Eaton, M. Czjzek, S. J. Perkins, T. P. Frandsen, M. SchÅElein, P. Panine, B. Henrissat, V. Receveur-Br echot, Protein disorder: Conformational distribution of the flexible linker in a chimeric double cellulase, *Biophys J* 88 (4) (2005) 2823 – 2832. doi:10.1529/biophysj.104.050146.
- [108] D. M. Ruiz, V. R. Turowski, M. T. Murakami, Effects of the linker region on the structure and function of modular gh5 cellulases, *Sci Rep* 6 (2016) 28504. doi:10.1038/srep28504.
- [109] B. Pau, M. Kristofer, G. Przemysław, S. Dmitri I., T. Marek, P. Miquel, M. Akke, Structure and dynamics of ribosomal protein L12: An ensemble model based on SAXS and NMR relaxation, *Biophys J* 98 (10) (2012) 2374–2382. doi:10.1016/j.bpj.2010.02.012.
- [110] D. W. Sammond, C. M. Payne, R. Brunecky, M. E. Himmel, M. F. Crowley, G. T. Beckham, Cellulase linkers are optimized based on domain type and function: Insights from sequence analysis, biophysical measurements, and molecular simulation, *PloS One* 7 (11) (2012) 1–14. doi:10.1371/journal.pone.0048615.
- [111] C. M. Payne, M. G. Resch, L. Chen, M. F. Crowley, M. E. Himmel, L. E. Taylor, M. Sandgren, J. Ståhlberg, I. Stals, Z. Tan, G. T. Beckham, Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose, *Proc Natl Acad Sci USA* 110 (36) (2013) 14646–14651. doi:10.1073/pnas.1309106110.
- [112] Szabo, Horvath, Schad, Murvai, Tantos, Kalmar, Chemes, Han, Tompa, Intrinsically disordered linkers impart processivity on enzymes by spatial confinement of binding domains, *Int J Mol Sci* 20 (9) (2019) 2119. doi:10.3390/ijms20092119.
- [113] N. E. Davey, The functional importance of structure in unstructured protein regions, *Curr Opin Struct Biol* 56 (2019) 155–163, sequences and Topology Carbohydrates. doi:10.1016/j.sbi.2019.03.009.

- [114] T. Flock, R. J. Weatheritt, N. S. Latysheva, M. M. Babu, Controlling entropy to tune the functions of intrinsically disordered regions, *Curr Opin Struct Biol* 26 (2014) 62–72, new constructs and expression of proteins / Sequences and topology. doi:10.1016/j.sbi.2014.05.007.
- [115] H. Wu, Higher-order assemblies in a new paradigm of signal transduction, *Cell* 153 (2013) 287–292. doi:10.1016/j.cell.2013.03.013.
- [116] N. E. Davey, G. Travé, T. J. Gibson, How viruses hijack cell regulation, *Trends Biochem Sci* 36 (3) (2011) 159–169. doi:10.1016/j.tibs.2010.10.002.
- [117] C. Charlier, G. Bouvignies, P. Pelupessy, A. Walrant, R. Marquant, M. Kozlov, P. De Ioannes, N. Bolik-Coulon, S. Sagan, P. Cortes, A. K. Aggarwal, L. Carlier, F. Ferrage, Structure and dynamics of an intrinsically disordered protein region that partially folds upon binding by chemical-exchange NMR, *J Am Chem Soc* 139 (35) (2017) 12219–12227. doi:10.1021/jacs.7b05823.
- [118] J. M. Krieger, G. Fusco, M. Lewitzky, P. C. Simister, J. Marchant, C. Camilloni, S. M. Feller, A. De Simone, Conformational recognition of an intrinsically disordered protein, *Biophys J* 106 (8) (2014) 1771–1779. doi:10.1016/j.bpj.2014.03.004.
- [119] R. E. Ithuralde, A. E. Roitberg, A. G. Turjanski, Structured and unstructured binding of an intrinsically disordered protein as revealed by atomistic simulations, *J Am Chem Soc* 138 (28) (2016) 8742–8751. doi:10.1021/jacs.6b02016.
- [120] G. Zhou, G. A. Pantelopulos, S. Mukherjee, V. A. Voelz, Bridging microscopic and macroscopic mechanisms of p53-MDM2 binding with kinetic network models, *Biophys J* 113 (4) (2017) 785–793. doi:10.1016/j.bpj.2017.07.009.

- [121] P. Robustelli, S. Piana, D. E. Shaw, D. E. Shaw, Mechanism of coupled folding-upon-binding of an intrinsically disordered protein, *J Am Chem Soc* 142 (25) (2020) 11092–11101. doi:10.1021/jacs.0c03217.
- [122] F. Paul, F. Noé, T. R. Weikl, Identifying conformational-selection and induced-fit aspects in the binding-induced folding of PMI from Markov state modeling of atomistic simulations, *J Phys Chem B* 122 (21) (2018) 5649–5656. doi:10.1021/acs.jpcc.7b12146.
- [123] P. Barthe, C. Roumestand, M. J. Canova, L. Kremer, C. Hurard, V. Molle, M. Cohen-Gonsaud, Dynamic and structural characterization of a bacterial FHA protein reveals a new autoinhibition mechanism, *Structure* 17 (4) (2009) 568–578. doi:10.1016/j.str.2009.02.012.
- [124] M. Maffei, M. Arbesú, A.-L. Le Roux, I. Amata, S. Roche, M. Pons, The SH3 domain acts as a scaffold for the n-terminal intrinsically disordered regions of c-Src, *Structure* 23 (2015) 893–902. doi:10.1016/j.str.2015.03.009.
- [125] M. Arbesú, M. Maffei, T. N. Cordeiro, J. M. C. Teixeira, Y. Pérez, P. Bernadó, S. Roche, M. Pons, The unique domain forms a fuzzy intramolecular complex in Src family kinases, *Structure* 25 (2017) 630–640. doi:10.1016/j.str.2017.02.011.
- [126] M. Arbesú, G. Iruela, H. Fuentes, J. M. C. Teixeira, M. Pons, Intramolecular fuzzy interactions involving intrinsically disordered domains, *Front Mol Biosci* 5 (2018) 39. doi:10.3389/fmolb.2018.00039.
- [127] K. Kasahara, M. Shiina, J. Higo, K. Ogata, H. Nakamura, Phosphorylation of an intrinsically disordered region of Ets1 shifts a multi-modal interaction ensemble to an auto-inhibitory state, *Nucleic Acids Res* 46 (5) (2018) 2243–2251. doi:10.1093/nar/gkx1297.
- [128] P. Klein, T. Pawson, M. Tyers, Mathematical modeling suggests cooperative interactions between a disordered polyvalent ligand and a single

- receptor site, *Curr Biol* 13 (2003) 1669–1678. doi:10.1016/j.cub.2003.09.027.
- [129] J. W. Locasale, Allovalency revisited: An analysis of multisite phosphorylation and substrate rebinding, *J Chem Phys* 128 (11) (2008) 115106. doi:10.1063/1.2841124.
- [130] X. Tang, S. Orlicky, T. Mittag, V. Csizmok, T. Pawson, J. D. Forman-Kay, F. Sicheri, M. Tyers, Composite low affinity interactions dictate recognition of the cyclin-dependent kinase inhibitor Sic1 by the SCFCdc4 ubiquitin ligase, *Proc Natl Acad Sci USA* 109 (9) (2012) 3287–3292. doi:10.1073/pnas.1116455109.
- [131] T. Mittag, S. Orlicky, W.-Y. Choy, X. Tang, H. Lin, F. Sicheri, L. E. Kay, M. Tyers, J. D. Forman-Kay, Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor, *Proc Natl Acad Sci USA* 105 (46) (2008) 17772–17777. doi:10.1073/pnas.0809222105.
- [132] M. Borg, T. Mittag, T. Pawson, M. Tyers, J. D. Forman-Kay, H. S. Chan, Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity, *Proc Natl Acad Sci USA* 104 (23) (2007) 9650–9655. doi:10.1073/pnas.0702580104.
- [133] T. Mittag, J. Marsh, A. Grishaev, S. Orlicky, H. Lin, F. Sicheri, M. Tyers, J. D. Forman-Kay, Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase, *Structure* 18 (2010) 494–506. doi:10.1016/j.str.2010.01.020.
- [134] S. Milles, D. Mercadante, I. Aramburu, M. Jensen, N. Banterle, C. Koehler, S. Tyagi, J. Clarke, S. Shammass, M. Blackledge, F. Gräter, E. Lemke, Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors, *Cell* 163 (3) (2015) 734–745. doi:10.1016/j.cell.2015.09.047.

- [135] R. Bayliss, T. Littlewood, M. Stewart, Structural basis for the interaction between fxfg nucleoporin repeats and importin- in nuclear trafficking, *Cell* 102 (1) (2000) 99–108. doi:10.1016/S0092-8674(00)00014-3.
- [136] T. N. Cordeiro, N. Sibille, P. Germain, P. Barthe, A. Boulahtouf, F. Allemand, R. Bailly, V. Vivat, C. Ebel, A. Barducci, W. Bourguet, A. le Maire, P. Bernadó, Interplay of protein disorder in retinoic acid receptor heterodimer and its corepressor regulates gene expression, *Structure* 27 (8) (2019) 1270–1285. doi:10.1016/j.str.2019.05.001.
- [137] I. M. S. de Vera, J. Zheng, S. Novick, J. Shang, T. S. Hughes, R. Brust, P. Munoz-Tello, W. J. Gardner, D. P. Marciano, X. Kong, P. R. Griffin, D. J. Kojetin, Synergistic regulation of coregulator/nuclear receptor interaction by ligand and DNA, *Structure* 25 (10) (2017) 1506–1518.e4. doi:10.1016/j.str.2017.07.019.
- [138] N. Rochel, F. Ciesielski, J. Godet, E. Moman, M. Roessle, C. Peluso-Iltis, M. Moulin, M. Haertlein, P. Callow, Y. Mély, D. I. Svergun, M. Dino, Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings, *Nat Struct Mol Biol* 18 (2011) 564–570. doi:10.1038/nsmb.2054.
- [139] L. Senicourt, A. le Maire, F. Allemand, J. E. Carvalho, L. Guee, P. Germain, M. Schubert, P. Bernadó, W. Bourguet, N. Sibille, Structural insights into the interaction of the intrinsically disordered co-activator TIF2 with retinoic acid receptor heterodimer (RXR/RAR), *J Mol Biol* (2021) 166899doi:10.1016/j.jmb.2021.166899.
- [140] E. F. Dudás, G. Pálffy, D. K. Menyhárd, F. Sebák, P. Ecsédi, L. Nyitray, A. Bodor, Tumor-suppressor p53TAD(1-60) forms a fuzzy complex with metastasis-associated S100A4: Structural insights and dynamics by an NMR/MD approach, *ChemBioChem* 21 (21) (2020) 3087–3095. doi:10.1002/cbic.202000348.

- [141] A. Shehu, L. E. Kavrakı, Modeling structures and motions of loops in protein molecules, *Entropy* 14 (12) (2012) 252–290. doi:10.3390/e14020252.
- [142] K. Kundert, T. Kortemme, Computational design of structured loops for new protein functions, *Biol Chem* 400 (3) (2019) 275–288. doi:10.1515/hsz-2018-0348.
- [143] A. Barozet, K. Molloy, M. Vaisset, T. Siméon, J. Cortés, A reinforcement-learning-based approach to enhance exhaustive protein loop sampling, *Bioinformatics* 36 (4) (2020) 1099–1106. doi:10.1093/bioinformatics/btz684.
- [144] W. Wang, D. Wang, Extreme fuzziness: Direct interactions between two IDPs, *Biomolecules* 9 (3) (2019). doi:10.3390/biom9030081.
- [145] J. Gsponer, M. Vendruscolo, Theoretical approaches to protein aggregation, *Protein Pept Lett* 13 (3) (2006) 287–293. doi:10.2174/092986606775338407.
- [146] T. P. J. Knowles, M. Vendruscolo, C. M. Dobson, The amyloid state and its association with protein misfolding diseases, *Nat Rev Mol Cell Biol* 15 (6) (2014) 384–396. doi:10.1038/nrm3810.
- [147] I. M. Ilie, A. Caffisch, Simulation studies of amyloidogenic polypeptides and their aggregates, *Chem Rev* 119 (12) (2019) 6956–6993, PMID: 30973229. doi:10.1021/acs.chemrev.8b00731.
- [148] S. Wu, D. Wang, J. Liu, Y. Feng, J. Weng, Y. Li, X. Gao, J. Liu, W. Wang, The dynamic multisite interactions between two intrinsically disordered proteins, *Angew Chem Int Ed* 56 (26) (2017) 7515–7519. doi:10.1002/anie.201701883.
- [149] A. Sottini, A. Borgia, M. Borgia, K. Bugge, D. Nettels, A. Chowdhury, P. Heidarsson, F. Zosel, R. Best, B. B. Kragelund, B. Schuler, Polyelectrolyte interactions enable rapid association and dissociation in

- high-affinity disordered protein complexes, *Nat Commun* 11 (2020) 5736. doi:10.1038/s41467-020-18859-x.
- [150] C. P. Brangwynne, P. Tompa, R. V. Pappu, Polymer physics of intracellular phase transitions, *Nat Phys* 11 (11) (2015) 899–904. doi:10.1038/nphys3532.
- [151] S. Boeynaems, S. Alberti, N. L. Fawzi, T. Mittag, M. Polymenidou, F. Rousseau, J. Schymkowitz, J. Shorter, B. Wolozin, L. Van Den Bosch, P. Tompa, M. Fuxreiter, Protein phase separation: A new phase in cell biology, *Trends Cell Biol* 28 (6) (2018) 420–435. doi:10.1016/j.tcb.2018.02.004.
- [152] Y.-H. Lin, J. D. Forman-Kay, H. S. Chan, Theories for sequence-dependent phase behaviors of biomolecular condensates, *Biochemistry* 57 (17) (2018) 2499–2508. doi:10.1021/acs.biochem.8b00058.
- [153] J. Wang, J.-M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, I. Poser, R. V. Pappu, S. Alberti, A. A. Hyman, A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins, *Cell* 174 (3) (2018) 688–699.e16. doi:10.1016/j.cell.2018.06.006.
- [154] E. W. Martin, A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, T. Mittag, Valence and patterning of aromatic residues determine the phase behavior of prion-like domains, *Science* 367 (6478) (2020) 694–699. doi:10.1126/science.aaw8653.
- [155] Y. C. Kim, G. Hummer, Coarse-grained models for simulations of multi-protein complexes: Application to ubiquitin binding, *J Mol Biol* 375 (5) (2008) 1416–1433. doi:10.1016/j.jmb.2007.11.063.
- [156] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, J. Mittal, Sequence

- determinants of protein phase behavior from a coarse-grained model, *PLoS Comput Biol* 14 (1) (2018) 1–23. doi:10.1371/journal.pcbi.1005941.
- [157] S. Das, Y.-H. Lin, R. M. Vernon, J. D. Forman-Kay, H. S. Chan, Comparative roles of charge, π , and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins, *Proc Natl Acad Sci USA* 117 (46) (2020) 28795–28805. doi:10.1073/pnas.2008122117.
- [158] M. K. Hazra, Y. Levy, Biophysics of phase separation of disordered proteins is governed by balance between short- and long-range interactions, *J Phys Chem B* 125 (9) (2021) 2202–2211, PMID: 33629837. doi:10.1021/acs.jpcc.0c09975.
- [159] A. E. Conicella, G. L. Dignon, G. H. Zerze, H. B. Schmidt, A. M. D’Ordine, Y. C. Kim, R. Rohatgi, Y. M. Ayala, J. Mittal, N. L. Fawzi, TDP-43 α -helical structure tunes liquid-liquid phase separation and function, *Proc Natl Acad Sci USA* 117 (11) (2020) 5883–5894. doi:10.1073/pnas.1912055117.
- [160] Z. Monahan, V. H. Ryan, A. M. Janke, K. A. Burke, S. N. Rhoads, G. H. Zerze, R. O’Meally, G. L. Dignon, A. E. Conicella, W. Zheng, R. B. Best, R. N. Cole, J. Mittal, F. Shewmaker, N. L. Fawzi, Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity, *EMBO J* 36 (20) (2017) 2951–2967. doi:10.15252/embj.201696394.
- [161] G. L. Dignon, W. Zheng, Y. C. Kim, J. Mittal, Temperature-controlled liquid-liquid phase separation of disordered proteins, *ACS Cent Sci* 5 (5) (2019) 821–830. doi:10.1021/acscentsci.9b00102.
- [162] R. M. Regy, G. L. Dignon, W. Zheng, Y. C. Kim, J. Mittal, Sequence dependent phase separation of protein-polynucleotide mixtures elucidated using molecular simulations, *Nucleic Acids Res* 48 (22) (2020) 12593–12603. doi:10.1093/nar/gkaa1099.

- [163] Z. Benayad, S. von Bülow, L. S. Stelzl, G. Hummer, Simulation of fus protein condensates with an adapted coarse-grained model, *J Chem Theory Comput* 17 (1) (2021) 525–537, PMID: 33307683. doi:10.1021/acs.jctc.0c01064.
- [164] T. S. Harmon, A. S. Holehouse, M. K. Rosen, R. V. Pappu, Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins, *eLife* 6 (2017) e30294. doi:10.7554/eLife.30294.
- [165] J.-M. Choi, F. Dar, R. V. Pappu, LASSI: A lattice model for simulating phase transitions of multivalent proteins, *PLoS Comput Biol* 15 (10) (2019) e1007028.
- [166] V. Nguemaha, H.-X. Zhou, Liquid-liquid phase separation of patchy particles illuminates diverse effects of regulatory components on protein droplet formation, *Sci Rep* 8 (1) (2018) 6728. doi:10.1038/s41598-018-25132-1.
- [167] J. R. Espinosa, J. A. Joseph, I. Sanchez-Burgos, A. Garaizar, D. Frenkel, R. Collepardo-Guevara, Liquid network connectivity regulates the stability and composition of biomolecular condensates with many components, *Proc Natl Acad Sci USA* 117 (24) (2020) 13238–13247. doi:10.1073/pnas.1917569117.
- [168] M. Paloni, R. Bailly, L. Ciandrini, A. Barducci, Unraveling molecular interactions in liquid–liquid phase separation of disordered proteins by atomistic simulations, *J Phys Chem B* 124 (41) (2020) 9009–9016. doi:10.1021/acs.jpcc.0c06288.
- [169] W. Zheng, G. L. Dignon, N. Jovic, X. Xu, R. M. Regy, N. L. Fawzi, Y. C. Kim, R. B. Best, J. Mittal, Molecular details of protein condensates probed by microsecond long atomistic simulations, *J Phys Chem B* 124 (51) (2020) 11671–11679, PMID: 33302617. doi:10.1021/acs.jpcc.0c10489.

- [170] S. Ambadipudi, M. Zweckstetter, Targeting intrinsically disordered proteins in rational drug discovery, *Expert Opin Drug Discov* 11 (1) (2016) 65–77, PMID: 26549326. doi:10.1517/17460441.2016.1107041.
- [171] A. Hatos, B. Hajdu-Soltész, A. M. Monzon, N. Palopoli, L. Álvarez, B. Aykac-Fas, C. Bassot, G. I. Benítez, M. Bevilacqua, A. Chasapi, L. Chemes, N. E. Davey, R. Davidović, A. Dunker, A. Elofsson, J. Gobeill, N. S. Foutel, G. Sudha, M. Guharoy, T. Horvath, V. Iglesias, A. V. Kajava, O. P. Kovacs, J. Lamb, M. Lambrugh, T. Lazar, J. Y. Leclercq, E. Leonardi, S. Macedo-Ribeiro, M. Macossay-Castillo, E. Maiani, J. A. Manso, C. Marino-Buslje, E. Martínez-Pérez, B. Mészáros, I. Mičetić, G. Minervini, N. Murvai, M. Necci, C. A. Ouzounis, M. Pajkos, L. Paladin, R. Pancsa, E. Papaleo, G. Parisi, E. Pasche, P. J. Barbosa Pereira, V. J. Promponas, J. Pujols, F. Quaglia, P. Ruch, M. Salvatore, E. Schad, B. Szabo, T. Szaniszló, S. Tamana, A. Tantos, N. Veljkovic, S. Ventura, W. Vranken, Z. Dosztányi, P. Tompa, S. C. E. Tosatto, D. Piovesan, DisProt: intrinsic protein disorder annotation in 2020, *Nucleic Acids Res* 48 (D1) (2019) D269–D276. doi:10.1093/nar/gkz975.