July 2019

Multiscale Simulations of Intrinsically Disordered Proteins

Xiaorong Liu

Multiscale Simulations of Intrinsically Disordered Proteins

A Dissertation Presented

by

XIAORONG LIU

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2019

Chemistry

Multiscale Simulations of Intrinsically Disordered Proteins

A Dissertation Presented

by

XIAORONG LIU

Approved as to style and content by:

_____
Jianhan Chen, Chair

_____
Scott Auerbach, Member

_____
Craig Martin, Member

_____
Li-Jun Ma, Member

_____
Richard Vachet, Department Head
Department of Chemistry

DEDICATION

To my parents, husband, advisor and other educators who light the way forward for me.

ACKNOWLEDGMENTS

First and foremost, I would like to express my wholehearted gratitude to my research advisor Dr. Jianhan Chen. Dr. Chen is a great scientist, who is extremely knowledgeable, wise and passionate about science. His dedication, open mindedness, and critical thinking greatly impact us. He is also a fantastic mentor, willing to share his knowledge, insights and experience with us, and incredibly supportive in our personal and career development. My graduate school journey couldn't be so rewarding and happy without his patient guidance, generous support, and continuous encouragement. I couldn't imagine having a better advisor for my PhD study, and it's my great honor to be his student.

I am also very grateful to other members in my dissertation committee, Dr. Scott Auerbach, Dr. Craig Martin and Dr. Li-Jun Ma. Their insightful comments and great suggestions always inspire me. I also want to thank our collaborators, Dr. Michal Zolkiewski and Dr. Om Prakash and Dr. Chungwen Liang for all of their help and encouragement.

I'm blessed to have had the opportunity of working in a wonderful research group. Many thanks to my labmates in Dr. Chen's lab: Dr. Zhiguang Jia, Dr. Kuo Hao Lee, Dr. Mara Chiricotto, Dr. Charles English, Dr. Azar Farjamnia, Dr. Weihong Zhang, Dr. Debabani Ganguly, Mahdieh Yazdani, Xiping Gong, Erik Nordquist, Lynn Schrag, Alex Beugelsdijk, Katrina Nguyen, Justin Campbell and Minh Ho. Their friendship, generous help, and stimulating discussion are greatly appreciated.

Last but not least, I am deeply indebted to my parents and my husband for their support and sacrifice. Their unconditional love, meticulous care, and full trust have nourished my life and encouraged me to move forward.

ABSTRACT

MULTISCALE SIMULATIONS OF INTRINSICALLY DISORDERED PROTEINS

MAY 2019

XIAORONG LIU

B.S., WUHAN UNIVERSITY, CHINA

M.S., WUHAN UNIVERSITY, CHINA

M.A., UNIVERSITY OF DELAWARE

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Jianhan Chen

Intrinsically disordered proteins (IDPs) lack stable secondary and/or tertiary structures under physiological conditions. The have now been recognized to play important roles in numerous biological processes, particularly cellular signaling and regulation. Mutation of IDPs are frequently associated with human diseases, such as cancers and neuron degenerative diseases. Therefore, it is important to understand the structure, dynamics, and interactions of IDPs, so as to establish the mechanistic basis of how intrinsic disorder mediates versatile functions and how such mechanisms may fail in human diseases. However, the heterogeneous structural ensembles of IDPs are not amenable to high resolution characterization solely through experimental measurements, and molecular modelling and simulation are required to study IDP structures, dynamics, and interactions at the atomistic levels.

Here, we first applied the state-of-the-art explicit solvent atomistic simulations to an anti-apoptotic protein Bcl-xL and demonstrated how inherent structural disorder may provide a

physical basis of protein regulated unfolding in signaling transduction. We have also constructed a series of efficient coarse-grained models to directly simulate the interactions between IDPs and unveiled how the preexisting structural elements accelerate binding of ACTR to NCBD by promoting efficient folding upon encounter. These studies shed important light on how IDPs perform functions in the cellular regulatory network, but also reveal the necessity of new sampling techniques for more efficient simulations of IDPs.

We have thus developed a novel sampling technique, called multiscale enhanced sampling (MSES). MSES couples the atomistic model with coarse-grained ones, to accelerate the sampling of atomistic conformational space. Bias from coupling to a coarse-grained model can be removed using Hamiltonian replica exchange. To achieve the best possible efficiency of MSES simulations, we have developed a new hybrid resolution protein model that could capture the essential features of IDP structures, so as to generate local and long-range fluctuations that are largely consistent with those at the atomistic level. We have also developed an advanced replica exchange protocol, to allow the fast conformational transitions observed in the coupled conditions to be rapidly exchanged to the unbiased limit. Application of these strategies to characterize the structural ensembles of a few non-trivial IDPs shows that faster convergence rate can be achieved, demonstrating the great potential of MSES for atomistic simulations of larger and more complex IDPs.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

1.1 Intrinsically disordered proteins: structure, dynamics, and interaction

Intrinsically disordered proteins (IDPs) lack well-defined three-dimensional (3D) structures under physiological conditions[1-7]. Although traditional protein structure-function paradigm suggests that proteins could perform functions due to their specific 3D structure, the discovery of IDP deviates from this notion and unveils the critical role of intrinsic disorder in mediating protein functions. The presence of functional disordered proteins, like the polypeptide hormone glucagon[8], was realized several decades ago, but the prevalence and versatile functions of IDPs were recognized only recently due to the availability of large amounts of gene sequence data, which enables us to perform gene-based functional analysis and to predict intrinsically disordered segments[9, 10]. Compared with folded proteins, IDPs are enriched in polar and charged residues, and deficient in bulky hydrophobic residues[11, 12]. By capturing such features of IDPs, sequence analysis have shown that about one third of eukaryotic proteins contain substantial disordered regions, and such disorder appears to be further enriched in proteins involved in cellular signaling and regulation[13]. The prevalence of IDPs suggests that many proteins rely on intrinsic disorder to perform functions. Also, mutation of IDPs is often associated with human diseases, such as cancer, diabetes, neuron degenerative disease, and cardiovascular disease[14]. Therefore, there is a great need to understand how intrinsic disorder mediate functions and how such mechanism may fail in human diseases.

1.1.1 Key properties and functional advantages of IDPs

It has been suggested that intrinsic disorder provides several unique advantages for the functional roles of IDPs. For example, intrinsic disorder provides a robust, thermodynamic basis for optimized allosteric coupling[15]. Unlike folded proteins, IDP structures are marginally stable and heterogeneous. Such disordered structural ensemble is poised to respond sensitively to input signals, allowing rapid redistribution among various conformational states, thus performing versatile functions[16]. Also, IDPs could often undergo disorder-to-order transitions upon specific binding to their targets to carry out functions, and such process is called “coupled binding and folding” of IDPs. Due to the malleable nature of IDPs, they could specifically interact with many targets by adopting different conformations[17]. Therefore, IDPs frequently work as signaling hub in the complex protein-protein interaction network and are rich in sites for posttranslational modifications[18, 19]. Moreover, intrinsic disorder provides a robust mechanism to decouple specificity from binding affinity, i.e., to allow for specific interaction without tight binding[20]. Such weak binding (and fast dissociation rate)[20] appeared to be particularly important for signaling transduction, such that signals could rapidly switch upon ligand binding, post-translational modification, or changes in cellular environment [21]. Also, it has been newly recognized that due to the weak, multivalent interactions of IDPs, they often play important roles in liquid-liquid phase separation, such as the formation of membrane-less organelles or biomolecular condensates[22]. Such macromolecular assemblies could function as reserves of signaling components[21, 23].

1.1.2 Experimental methods towards characterizing IDP structures

One focus of IDP studies is to understand how the structural and dynamic properties of unbound IDPs impact their interactions and functions. As mentioned above, theoretical studies have predicted that intrinsic disorder provides a robust, thermodynamic basis for optimized allosteric regulation[15]. It has also been suggested that conformational flexibility of IDPs is advantageous for efficient molecular recognition, since it may allow IDPs to weakly interact with their targets at a larger distance[24]. In contrast, evidence also exists that pre-formed structural elements in the unbound ensemble of IDPs facilitate their coupled binding and folding, probably by providing initial binding interface[25, 26]. Furthermore, the structural and dynamic properties of IDPs can be modulated by posttranslational modifications[18] and changes in cellular environment[27], like temperature, pH, or ionic strength, thus altering their interaction and function. To gain better understanding of the underlying mechanisms, one of the most important requirements is detailed characterization IDP unbound ensembles. Experimentally, nuclear magnetic resonance (NMR)[28, 29], circular dichroism (CD)[30, 31], single-molecule Förster resonance energy transfer (sm-FRET)[32, 33], small angle X-ray (or neutron) scattering (SAXS or SANS)[34-37] have been widely used to characterize the structural properties of IDPs. For instance, NMR chemical shift and CD spectrometry are particularly useful to quantify the level of secondary structure in IDPs; NMR paramagnetic resonance enhancement and sm-FRET could provide distance information between different residues; SAXS and SANS are powerful tools in determining the overall dimension of IDPs; NMR nuclear Overhauser effect and relaxation studies are commonly used to study IDP dynamics. Many IDPs have been characterized using these techniques, such as p53[38-40],

α-synuclein[41-43], prions[33] and Aβ[44-47]. Many important structural features have been observed for IDPs, including less compactness compared with molten globule proteins, the presence of some residual secondary structure, and possible transient long-range contact formation.

1.1.3 Experimental methods for studying IDP interactions

Apart from characterizing structure and dynamics of unbound IDPs, direct mechanistic study of IDPs interactions is equally important for elucidating the molecular mechanisms of how IDPs perform functions. As mentioned earlier, IDPs could often fold into stable 3D structure upon specific binding to their physiological partners, and in many cases, such coupled binding and folding processes are associated with their activity and can be modulated by a lot of cellular events[48]. However, it's under debate whether the baseline mechanism is induced folding (i.e., binding occurs before folding) or conformational selection (i.e., folding occurs before binding). Also, it remains unclear how pre-formed structural element in the unbound ensemble of IDPs modulates the pathways of coupled binding and folding. More intricately, polymorphism and dynamics of the bound state has recently been observed[49-51], which offers another toolbox for regulation. But the driving forces of such dynamic recognition are yet to be understood. To gain mechanistic insights of how IDPs achieve both fidelity and efficiency in molecular recognition, many experimental techniques could be utilized to characterize IDP interactions, like NMR relaxation dispersion[52] and stop-flow fluorescence measurement[31]. Fitting of relaxation dispersion data to appropriate multiple-state models could unveil the presence possible intermediate states, and directly yield exchange rates between these states[53]. Stop-flow fluorescence measurement provides a means to obtain association and

dissociation rate constants. Thus, binding affinity could also be derived from the measurement. In combination of mutagenesis, these kinetic and thermodynamic information can be further analyzed through ϕ-value analysis and linear free energy relationship[54]. By calculating the ratio of activation free energy difference to binding free energy difference due to single-site mutation, ϕ-value analysis is a valuable tool to determine if a specific interaction is formed before or after transition state. In a similar spirit, linear free energy relationship could report the overall mechanism of IDP coupled and binding by analyzing the changes in activation free energy and binding free energy for a series of mutants.

1.1.4 Challenges in experimental studies of IDPs

Although a lot of important insights have been gained from existing studies, it's also well-recognized that formidable challenges exist for obtaining atomistic details of IDP structure, dynamics and interactions solely from experimental measurements. In many cases, only ensemble averaged quantities can be obtained from experimental studies, and how to recover the underlying distribution from ensemble averages is a severely under-determined problem[53]. Such heterogeneous nature of IDP ensembles could also complicate many analyses. For instance, the NMR relaxation dispersion data can only be practically fitted to two- or three-state models, and ambiguity in interpreting the derived kinetic parameters may arise if a complex IDP interaction involves additional intermediate states[53]. Similarly, reliable interpretation of ϕ-value analysis may not be feasible if non-native interactions are present in unbound state or transition state, or if there is any intermediate state during interaction. Although sm-FRET could provide time series of information instead of just ensemble averages, a key challenge is the need of covalent labelling using

fluorescent dyes, which may generate artifacts and limit the minimal detecting distance to at least 2 nm[55]. Besides the structural heterogeneity of IDPs, timescales of conformational fluctuations may further impose difficulties in analyzing experimental measurements. For example, NMR relaxation dispersion analysis can only be applied to slow exchange between conformational states (i.e., in microsecond to millisecond timescale). Therefore, alternative methods are required to complement and extend the understanding gained from experimental studies.

1.2 Molecule modeling and simulations in studying IDPs

These experimental challenges likely represent opportunities for molecular modelling and simulation to make unique and critical contributions to studying IDPs. Recent advances in computing hardware, force field and enhanced sampling methods have made it possible to reliably simulate IDP conformations. In principle, molecular dynamics (MD) simulations using the physics-based atomistic force fields could directly provide high-resolution spatial and temporal information about IDP structure and dynamics. Also, the design of powerful computer hardware, including distributed computing (e.g., Folding@home[56, 57]), graphics processing unit (GPU) and special-purpose computers (e.g., Anton[58]), have dramatically increased the simulation speed and allowed for much longer timescale of atomistic simulations. Moreover, many enhanced sampling methods have been developed such that less computation is required in order to generate converged ensembles.

1.2.1 Computational studies of IDP structure and dynamics

Molecular modelling and simulations have provided many important insights for deeper understanding of IDP structural and dynamical properties which are not amenable to experimental characterization. Among many atomistic approaches, the implicit solvent force fields, especially the GBSW model[59] and ABSINTH[60], have been successfully applied to detailed studies of many IDPs in their unbound state. Such implicit treatment of solvent could provide an appropriate balance between efficiency and accuracy. By careful optimization, the GBSW model has properly balanced solvation and intramolecular interactions[59]. But compared with explicit solvent simulations, GBSW simulations of small proteins are ~30 times faster, which makes it possible to extensively explore the vast conformational space available to IDPs. A particularly interesting example is the atomistic simulations of the kinase-inducible domain (KID) of transcription factor CREB using GBSW[61]. By comparing the structural ensembles of KID, its phosphorylated form (pKID) and a mutant S133E, it has been found that phosphorylation of S133 not only provides an interaction site for binding to its partner, the KIX domain of coactivator CBP (CREB-binding protein), but also reduces the entropic cost of KID folding upon binding[61]. Similar approaches have also been used to study another IDP, the transactivation domain (TAD) of tumor suppressor p53 and a few cancer-associated mutants[62]. These cancer mutations could modulate both local and long-range conformational properties of p53-TAD, which may further affect its binding to various regulatory proteins, thus altering the function. In parallel, the ABSINTH force field has been widely used to study the relationship between amino acid

sequence and polymeric properties of IDPs. For example, a series of arginine-rich protamines have been simulated using the ABSINTH model, which demonstrates that increasing net charge per residue could induce transitions of IDPs from compact globules to swollen coils[63]. Also, by studying 30 variants of $(EK)_{25}$ peptides, it has been found that the charge distribution of polyampholytic IDPs could modulate their conformational preference. Well-mixed sequences usually behave like random coils while segregation of opposite charges leads to hairpin-like compact conformations[64].

1.2.2 Computational studies of IDP interactions

There have been much less atomistic simulations to study the molecular mechanisms of IDP coupled binding and folding compared with studies of unbound IDPs, mainly because of the computational cost. A noteworthy example is the calculation of free energy profiles for the C-terminus of p53 binding to S100B(ββ) using atomistic simulations[65]. Implicit solvent model GBSW was used in combination with enhanced sampling methods including both temperature-replica exchange and umbrella sampling. This study reveals that p53 binds to S100B(ββ) in an unfolded state and then folds rapidly, i.e., following an induced folding mechanism. Many other studies use coarse-grained models to study IDP interactions, in order to reach the biological timescale of such processes (i.e., microseconds to milliseconds), and to obtain statistically meaningful observations by sampling numerous reversible binding/folding transitions. For instance, topology-based Gō or Gō-like models[66] are simple but very successful in studying the mechanisms of coupled binding and folding of IDPs. These models are based on the minimal frustration theory of protein

folding, which predicts that native interactions largely shape the funnel-like energy landscape[67, 68]. Therefore, given the topology of a folded protein, one could construct effective energy functions based on the native contacts to capture the key features of true energy landscape. In an example of a cell cycle regulator $p27^{Kip1}$ (p27) interacting with cyclin A, Gō model simulations have shown that long-range electrostatic interactions could not only increase the encounter rate, but also promote efficient folding upon encounter[69]. Similar approaches have been used to investigate the coupled binding and folding of pKID interacting with KIX[70]. Increasing the amount of residual structure in unbound pKID was found to decelerate KIX binding, suggesting that intrinsic disorder of IDPs is advantageous for target recognition.

1.2.3 Limitations in computational approaches

However, limitations still exist in previous simulations of IDP structure and interactions. For instance, the implicit solvent force fields assume that the highly dynamic solvent molecules can be treated as a continuum medium, which is problematic if non-bulk waters play important roles in interactions. Also, the GBSW model is inherently limited by its inaccurate description of protein-water boundaries. In coarse-grained approaches, atomistic details of IDP structures are lost. Specifically, the Gō models are inadequate to describe the heterogeneous structures of unbound IDPs, and difficult to examine the role of non-native interactions in IDP coupled binding and folding. Hence, there is a great need to develop more accurate force fields and more efficient algorithms to further improve the reliability and extend the applicability of computational simulations in studying IDPs.

1.3 Advances and challenges in simulating IDPs

1.3.1 Advances in force field development

Reasonable description of numerous sub-states available to IDPs and free energy barriers between them pushes the limit of force field accuracy. Previous generation of force fields were usually optimized using the native state of folded proteins, and often appeared to generate overly compact conformations for many IDPs[71]. Therefore, there have been extensive efforts towards improving the quality of protein force fields. As discussed above, implicit solvent force fields provide optimal balance between physical accuracy and computational cost, but the accuracy of GBSW model is limited by its inaccurate description of protein-water boundaries. A more appropriate dielectric boundary can lead to improved accuracy, just like in the generalized Born with molecular volume model (GBMV2). GBMV2 model recently has been optimized[72], which could capture the structure and stability of both α-helical and β-hairpin model peptides. Also, the simulated ensemble of several IDPs with various lengths are highly consistent with experimental observables. Explicit solvent all-atom force fields arguably offer more accurate representation of proteins, although careful optimization requires demanding computational resources to achieve sufficient sampling. Many exciting advancements have been made recently. For example, CHARMM36m[73] has been validated using a set of 15 peptides and 20 proteins. Improved accuracy of backbone conformations was observed for both folded proteins and IDPs, while overly compact structures were still found for larger IDPs. Force field a99SB-disp[71] represents another significant improvement in force field development, which has been benchmarked by a series

of peptides and proteins with various lengths and stabilities. This model could accurately describe folded proteins, secondary structure and overall dimension of disordered proteins. Many other force fields aiming at better description of IDPs can also be found, such as ABSINTH[60], ff99SB-ILDN[74, 75] with TIP4P-D water model[76] and ff03ws[77].

1.3.2 Development of enhanced sampling methods

To visit the vast conformational space available to IDPs also pushes the sampling capacity in MD simulations. One of the most popular enhanced sampling methods is temperature replica exchange (T-RE)[78], where multiple replicas of the same system are simulated under different temperatures, with their temperatures exchanged periodically according to the Metropolis criterion. The essence of T-RE is that higher temperature is effective in accelerating enthalpy-limited processes, thus facilitating the escape from local energy minima, speeding up conformational transitions and enhancing sampling. T-RE has been very successful in studying biomolecules including IDPs[61, 79-82]. However, it's also relatively well known that limitations exist in studying large-scale conformational transitions. For example, replicas are often trapped in conformational space, and different replicas tend to sample different conformations[83]. Such segregation in conformational space may hinder the random walk of replicas in temperature space and lead to exchange bottlenecks near the melting temperature where large transitions occur. Many strategies have been explored in order to alleviate these problems. For example, to obtain uniform exchange acceptance probability has been suggested to be important for enhanced sampling, which can be achieved through concentrating

replicas near the temperatures of exchange bottleneck[84]. However, only considering exchange probability may not necessarily guarantee improved sampling efficiency. A more sophisticated approach is to optimize the flow of replicas in temperature space, such as maximizing the rate of replica round trips between the lowest and highest temperatures by adaptively adjusting the distribution of simulation temperatures[84, 85] or by using biasing potentials[86, 87], and steadily increasing or decreasing the temperature of a replica to help it escape from the trapped state[88]. Nevertheless, if the system needs to undergo slow, cooperative conformational transitions, like protein folding, where the activation free energy is entropically dominated, tempering likely becomes ineffective in driving such transitions.

Besides T-RE and its variants, many other enhanced sampling methods have also been developed to study protein conformations. Accelerated molecular dynamics method[89] enhances sampling by adding non-negative bias potential to the system where system potential is lower than a given value while keeping other regions, like transition states, unchanged. Similarly, if the location of a system during the simulation can be described by a few collective variables (CVs), metadynamics[90] can be used by continuously adding Gaussian potentials centered on the previously visited points in the CV space. The accumulation of such Gaussian bias potentials will flatten the free energy landscape and push the system to explore the whole conformational space. Moreover, instead of using Gaussian potentials, variational approach can be utilized to determine bias potential, such that the system can sample an arbitrary probability distribution in the CV space[91]. Another unique approach

is orthogonal space sampling[92], where sampling is enhanced in both the order parameter space and its generalized force space. The basic idea is that the generalized force is a robust order parameter to describe the direction along which hidden barriers exist. Therefore, this approach could accelerate random walking along the selected order parameter as well as its environmental response, thus enhancing conformational transitions.

The efficacy of these enhanced sampling methods in studying IDPs remains to be carefully examined. Since conformational transitions of IDPs between numerous sub-states cannot be easily described by a few CVs, it may complicate the evaluation of sampling efficiency of many approaches that depends critically on choosing appropriate CVs. Moreover, the free energy barriers of large conformational transitions, like protein folding, are usually dominated by the entropic component[93], which makes tempering ineffective in driving such transitions. In this dissertation, strategies towards alleviating such problems will be discussed.

1.4 Dissertation Outline

The structure, dynamics and interaction of IDPs is of great scientific interest due to their biological significance. This dissertation will include application of multiscale modelling and simulations to understand how intrinsic disorder mediates protein functions, as well as development of new enhanced sampling methods for more efficient simulation of IDPs. Specifically, in chapters 2[94] and 3[95], atomistic simulations were performed to study how an anti-apoptotic protein Bcl-xl use intrinsic disorder to mediate regulated unfolding in cell apoptosis. Chapter 4[96] is

focused on interaction between two IDPs, activation domain of the activator for thyroid hormone and retinoid receptors (ACTR) and the nuclear coactivator binding domain (NCBD) of CBP. A series of coarse-grained models were constructed to investigate the molecular mechanisms of how residual structure in ACTR accelerates binding with NCBD. These studies help us better understand how IDPs perform versatile functions, but again, suggest that new sampling techniques are needed for more efficient atomistic simulations of IDPs. Therefore, the next two chapters will be focused on strategies to improve a novel enhanced sampling technique, called multiscale enhanced sampling (MSES) where coarse-grained models are coupled with all-atom ones to accelerate sampling of atomistic energy landscape. The bias from coupling to a coarse-grained model could be removed by replica exchange, thus allowing us to benefit simultaneously from the high accuracy of atomistic models and faster dynamics of coarse-grained models. In particular, chapter 5[97] presents the development of a better coarse-grained model for IDPs, which could generate fluctuations largely consistent with those at the atomistic levels, thus leading to better synergy between atomistic and coarse-grained models and accelerating atomistic structural transitions. Chapter 6 shows the development of advanced replica exchange protocols for MSES, which allows for better control of how the coupled conditions are exchanged to unbiased limit. Finally, a brief summary of findings and directions for further research are discussed in chapter 7.

CHAPTER 2

DYNAMICS OF THE BH3-ONLY PROTEIN BINDING INTERFACE OF BCL-XL*

The balance and interplay between pro-death and pro-survival members of the B-cell lymphoma-2 (Bcl-2) family proteins play key roles in regulation of the mitochondrial pathway of programed cell death. Recent NMR and biochemical studies have revealed that binding of the pro-apoptotic BH3-only protein PUMA induces significant unfolding of anti-apoptotic Bcl-xL at the interface, which in turn disrupts the Bcl-xL/p53 interaction to activate apoptosis. However, the molecular mechanism of such regulated unfolding of Bcl-xL is not fully understood. Analysis of the existing PDB structures of Bcl-xL in both bound and unbound states reveal substantial intrinsic heterogeneity at its BH3-only protein binding interface. Large-scale atomistic simulations were performed in explicit solvent for six representative structures to further investigate the intrinsic conformational dynamics of Bcl-xL. The results support that the BH3-only protein binding interface of Bcl-xL is much more dynamic compared to the rest of the protein, both unbound and when bound to various BH3-only proteins. Such intrinsic interfacial conformational dynamics likely provides a physical basis that allows Bcl-xL to respond sensitively to detailed biophysical properties of the ligand. The ability of Bcl-xL to retain or even enhance dynamics at the interface in bound states could further facilitate the regulation of its

* Liu X, Beugelsdijk A, and Chen J (2015). "Dynamics of the BH3-Only Protein Binding Interface of Bcl-xL." Biophysical Journal 109(5): 1049-1057.

interactions with various BH3-only proteins such as through post-translational modifications.

2.1 Introduction

Intrinsically disordered proteins (IDPs) frequently play crucial roles in cell signaling and regulation and are associated with numerous human diseases [2, 3, 15, 17, 48, 53, 98]. Intrinsic conformational disorder of IDPs may offer many potential functional advantages, such as larger binding surface areas, inducibility by posttranslational modifications, and structural plasticity for binding multiple partners [3, 53]. Attesting to the fundamental importance of intrinsic disorder in biology, sequence analysis has revealed that over one-third of eukaryotic proteins contain long disordered segments or domains [99]. Intensive efforts have been focused on characterizing the conformational properties of unbound IDPs and understanding how these properties may support facile and robust binding to specific targets [100, 101]. The ability of many regulatory IDPs to undergo coupled binding and folding transitions upon specific binding, in particular, has attracted much attention [100, 102]. It is also increasingly recognized that substantial conformational heterogeneity, and sometimes full disorder of the entire binding domain, may persist in the bound states of IDPs [103-106]. Furthermore, examples have started to emerge in recent years where cellular signaling and regulation are achieved via regulated unfolding of proteins [107]. Regulated unfolding could be local or global, and may be driven by a wide range of signaling stimuli, including post-translational modifications [108], ligand or protein binding [109-111], changes

in environmental conditions such as pH [112, 113], and mechanical stress [114, 115]. Together, these examples illustrate a fascinatingly broad and versatile utilization of protein conformational disorder in cellular signaling and regulation.

A particularly intriguing example of regulated unfolding involves the binding of intrinsically disordered PUMA to protein Bcl-xL [109]. Bcl-xL is an anti-apoptotic member of the Bcl-2 family proteins, which are critical regulators of the mitochondrial pathway of programmed cell death [116]. Interactions and balance between pro-apoptotic and anti-apoptotic members of the Bcl-2 family proteins underlie the regulatory network that controls the switch between life and death of the cell. Mis-regulation of the Bcl-2 family proteins is frequently involved in cancers [117]. Bcl-xL inhibits the pro-apoptotic function of cytoplasmic tumor suppressor p53 by sequestering it into inactive complexes [118, 119]. It also protects cells from programmed death by interactions with numerous pro-apoptotic BH3-only Bcl-2 family proteins, including BID, BIM, BAD, PUMA, BIK, HRK, BMF and NOXA [120-122]. All BH3-only Bcl-2 family proteins except BID are IDPs [123]. Upon binding to Bcl-xL, the BH3 domain of PUMA folds into a single helix [109] (Figure 2.1B). Intriguingly, PUMA binding also leads to local unfolding of Bcl-xL [109], mainly in the α2 and α3 segments near the BH3-only protein binding interface (see Figure 2.1B), which in turn disrupts the interactions with cytosolic p53 and releases the inhibition of p53's pro-apoptotic function. The apparently unique ability of PUMA binding to drive drastic local unfolding of Bcl-xL appears consistent with the fact that PUMA is the only BH3-only protein that can efficiently release p53

from the inactive complex with Bcl-xL [124]. However, the molecular basis of regulated unfolding in the Bcl-xL/PUMA interaction is not fully understood.

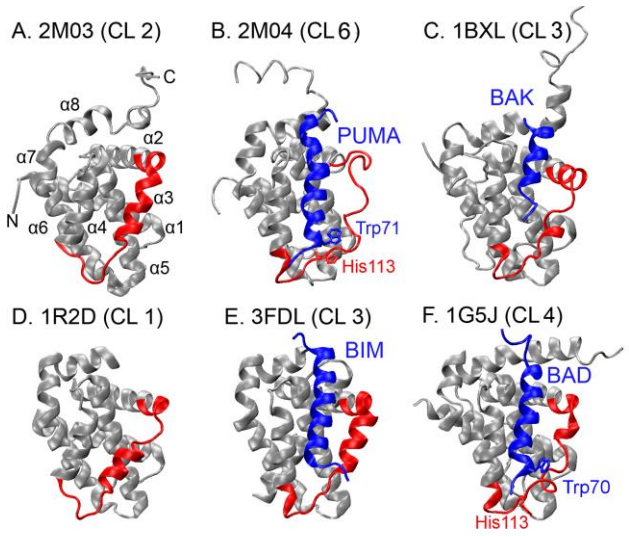


Figure 2.1 Representative experimental structures of the BH3-only protein binding interface of Bcl-xL. Bcl-xL is colored in gray except for residues 98-120, which are colored in red. The BH3-only protein binding partners (PUMA, BIM and BAD) are shown in blue. Additional details of these PDBs are provided in Table 2.2. The cluster IDs are shown in parenthesis (see Table 2.3 for details).

PUMA contains a unique Tryptophan at position 71 among BH3-only proteins. Importantly, its π-stacking interaction with Bcl-xL His113 has been shown to be necessary for the observed regulated unfolding of Bcl-xL [109] (Figure 2.1B). Mutation of PUMA Trp71 to Ala largely suppresses Bcl-xL partial unfolding in the bound state, and further abolishes the ability of PUMA binding to release cytosolic p53 from Bcl-xL inhibition [109]. These observations have led to the conclusion that the π-stacking interaction between PUMA Trp71 and Bcl-xL His113 drives local unfolding in the adjacent α2 and α3 segments of Bcl-xL. However, the π-stacking interaction itself does not appear to be thermodynamically important, because the W71A mutant and wild-type PUMA BH3 domain bind to Bcl-xL with similar high

affinities [109]. It appears unusual that a single specific interaction between Bcl-xL His113 and PUMA Trp71, located on the edge of the binding interface, could fully account for the dramatic conformational changes observed for Bcl-xL. Curiously, the BH3 domain of BAD also contains a Tryptophan at position 70 (PUMA numbering), which is in position to potentially make similar π-stacking interaction with Bcl-xL His113 (Figure 2.1F), but it does not lead to similar local unfolding upon binding [109] (we note that α3 does become sufficiently distorted in PDB:1G5J such that it is not fully assigned to the helical state in the secondary structure analysis; see Figure 2.1F and Figure 2.2). In this work, we analyze all existing Protein Data Bank structures of Bcl-xL both in the apo form and in complex with various small molecule and peptide ligands, and perform extensive molecular dynamics (MD) simulations for six selected Bcl-xL apo structures and complexes to characterize the inherent conformational dynamics of Bcl-xL. The results together demonstrate that that the BH3-only protein binding interface of Bcl-xL is much more dynamic than currently recognized. Such intrinsic interfacial conformational dynamics could provide a physical basis that enables Bcl-xL to respond sensitively to the nature of the bound ligand and/or environmental conditions, thus allowing facile unfolding upon specific binding of appropriate ligands such as PUMA.

2.2 Methods

A total of 45 PDB entries were identified that contain either unbound Bcl-xL monomer or its complex with small molecule or peptide ligands (as of September 2014) [109, 119, 125-145]. Four domain swapped dimers of Bcl-xL, namely, 3FDL,

2YQ6, 2YQ7 and 4A1U, are also included, where α1 is involved in intermolecular interactions and adopts a very different configuration (e.g., Figure 2.9). Information about all PDB entries included in the current analysis is provided in Table 2.2. For structural analysis, all entries were preprocessed to remove extra atoms that do not belong to Bcl-xL or its primary ligand (if present). The residue numbering of Bcl-xL was modified in all PDB entries to match the one used in 2M04 (Bcl-xL/PUMA) for convenience of comparison. For PDB entries containing multiple models/chains, only chain A of model 1 was included the analysis. Clustering analysis was performed using the fixed radius clustering algorithm as implemented in the MMTSB/cluster.pl tool (with -kclust option) [146].

Atomistic MD simulations in explicit solvent were performed to further characterize the structure and dynamics of Bcl-xL in both bound and unbound states. These simulations were initiated from six representative PDB structures. Two were based on the NMR and X-ray structures of unbound Bcl-xL, namely, 2M03 [109] and 1R2D [127]. The other four simulations were based on the complex structures of Bcl-xL with various BH3-only protein ligands, including 1BXL for Bcl-xL/BAK [134], 1G5J for Bcl-xL/BAD [147], 2M04 for Bcl-xL/PUMA [109], and 3FDL for Bcl-xL/BIM [148]. We note that the long and presumably disordered loop between α1 and α2 is absent in all original PDB structures noted above. To our best knowledge, this loop is not involved in the complex formation between BH3-only peptides and Bcl-xL, and thus omitted in the current simulations. A caveat is that the potential impacts of the disordered loop on the structure and dynamics of the rest of the protein will not be captured. All structures were solvated using TIP3P water

molecules, with proper counter ions added to neutralize the whole system. The final solvated systems are illustrated in Figure 2.10. They consist of 40,740 to 97,371 atoms with cubic box dimensions ranging from ~73 to 98 Å. Table 2.1 provides a summary of the residue ranges, total atom numbers, box sizes, and total simulation times of all simulations.

Each solvated system was energy minimized using steepest descent and adopted basis Newton-Raphson methods, followed by a short equilibration simulation of 100 ps using CHARMM [149, 150] with a small harmonic restraint imposed on protein heavy atoms to slowly relax the system. After that, unrestrained molecular dynamics production simulations were performed using NAMD [151]. CHARMM36/CMAP force field was used to model proteins, water and ions [152-154]. These simulations were carried out under constant temperature (298 K) and constant pressure (1 bar), and periodic boundary conditions were imposed. Non-bonded interactions were truncated at the distance of 13 Å, with a smooth switching function starting from 12 Å. The pair list was updated every 10 steps with a pair list distance of 15 Å. Long-range electrostatic interactions were treated with Particle Mesh Ewald (PME) method [155] with a grid spacing of ~1 Å. Lengths of all hydrogen-related bonds were kept constant using SHAKE algorithm [156], and a time step of 2 fs was used to integrate the equations of motion. The total production simulation lengths range from 310 ns (for 1BXL) to 730 ns (for 1R2D) (see Table 2.1). All analysis was performed using CHARMM and in-house scripts, and molecular illustrations were prepared using the visual molecular dynamics (VMD) software [157].

Table 2.1 Summary of all six simulations

PDB	Proteins	Bcl-xL Residues [a]	Water No.	Na^+ No.	Total Atom No.	Initial Box Size (Å)	Simulation Time (ns)
1R2D	Bcl-xL	1-27, 82-196	12824	6	40740	73.0	730
2M03	Bcl-xL	1-44, 85-200	13660	12	43498	74.8	570
2M04	Bcl-xL/PUMA	1-44, 85-200	18237	14	57654	82.4	460
1G5J	Bcl-xL/BAD	1-44, 85-211	14233	10	45836	76.5	580
3FDL	Bcl-xL/BIM	1-26, 83-194	17922	5	56421	81.6	450
1BXL	Bcl-xL/BAK	1-44, 85-217	31439	12	97371	98.1	310

[a] Residue numbering is based on that of 2M04

2.3 Results and Discussion

2.3.1 Analysis of existing PDB structures of Bcl-xL

We first analyzed existing PDB structures of Bcl-xL in both bound and unbound states to examine the conformational flexibility of the BH3-only protein binding interface of Bcl-xL. Bcl-xL structure consists of eight helices (α1-8) connected by loops of different lengths (Figure 2.1A). As summarized in Figure 2.2, most helices are consistently present in all PDB structures except α3. Interestingly, PUMA does not appear to be the only ligand that can drive partial unfolding in the α2/α3 region of Bcl-xL. Instead, α3 appears to be unfolded in many complexes, such as 3ZK6, 3ZLN, 3ZLO, 2O2N, 1YSG, 1YSI, 1BXL, 1G5J, 3PL7 and 4A1U. These complexes involve both small molecule and peptide ligands, none of them contains analogous π-stacking interactions to the one between PUMA Trp71 and Bcl-xL His113 that has been postulated to be critical in driving partial unfolding in Bcl-xL (except potentially for 1G5J; see Figure 2.1F). Even for the structures where α3 is not fully unfolded, the length and position of α3 can vary significantly among different PDB entries. For example, in 2LPC, α3 consists of 10 residues (residues 102-111), while

in 1PQ1, α3 is only composed of 4 residues (residues 108-111). Other regions near the BH3-only protein binding interface, including both C-terminus of α2 and N-terminus of α4, also display substantial variance among different PDB structures. In particular, the C-terminus of α2 becomes disordered in two apo structures of Bcl-xL (1LXL and 2ME9), similar to what was observed in the Bcl-xL/PUMA complex (2M04).

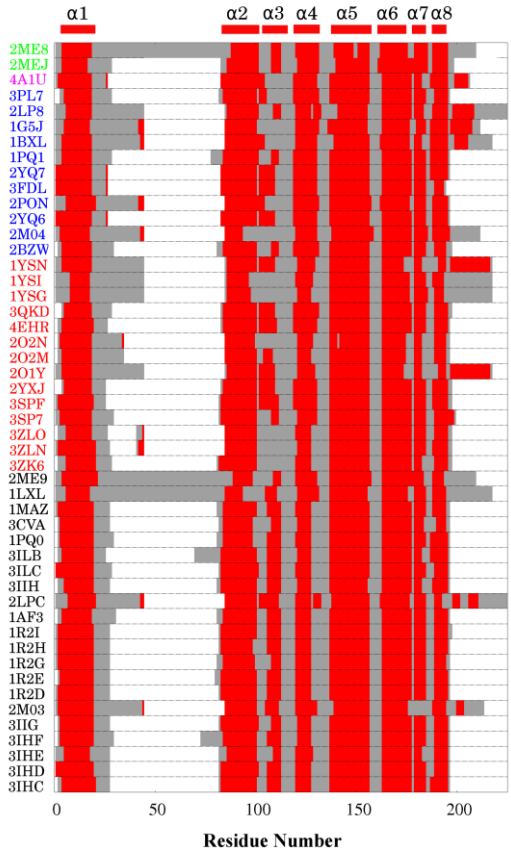


Figure 2.2 Secondary structures of Bcl-xL in 49 PDB entries in 2M04 numbering scheme. Helical residues are colored in red and disordered/loop residues in gray. White regions mark residues absent in the PDB structure. The PDB entries are ordered and colored based on the nature of the bound ligands: p53 (green), an α-β-foldamer homologues of the BIM BH3 domain (magenta), BH3-only peptides/proteins (blue), small molecules (red), and unbound (black). Note that some structures contain single or double mutations in Bcl-xL (see Table 2.2 for details).

Variations in the Bcl-xL PDB structures can be quantified by calculating the root-mean-square-fluctuation (RMSF) profiles. RMSF quantifies the magnitude of atomic positional fluctuation around the mean, and has been shown to correlate strongly with order parameters derived from NMR relaxation analysis [158]. For this, all structures were first aligned using C_α atoms in the core region, which was identified as regions with minimal secondary and tertiary structure variations and included residues 85-98, 123-127, 140-156 and 162-175. We note that RMSF profiles calculated with structures aligned using the whole protein are not qualitatively different (e.g., see Figure 2.11). Bcl-xL in domain swapped-dimers contains a dramatically different pose of α1 (see Figure 2.9). Inclusion of the four domain-swapped dimer PDBs thus leads to artificially high RMSF values in α1 (Figure 2.11 black trace). Excluding these four PDBs yields a RMSF profile that more accurately reflects the structural variations within the Bcl-xL monomer in bound and unbound states (Figure 2.11 red trace). Clearly, the BH3-only protein binding interface of Bcl-xL is highly variant. The RMSF values exceed 3.5 Å in regions near α3, and are much higher than most loops except the long one between α1 and α2 (which is disordered and not resolved in most structures). Importantly, these variations among PDB structures seem to reflect the intrinsic conformational dynamics of Bcl-xL. For example, the RMSF profile converted from the B-factors of a representative X-ray crystal structure of unbound Bcl-xL (1R2G), shown as the green trace in Figure 2.3, is highly consistent with that derived from all PDBs (red trace), which also suggest that regions near α3 is much more dynamic than the rest of the protein. Similar observations can be made on the RMSF profile calculated

from the 20-member NMR ensemble of 2M04 (Figure 2.12). The notion that the BH3-only protein binding interface of Bcl-xL is intrinsically more dynamic is further supported by sequence analysis by PrDOS [159], which predicts elevated disorder propensities near α3/4 compared the other helical regions (Figure 2.13).

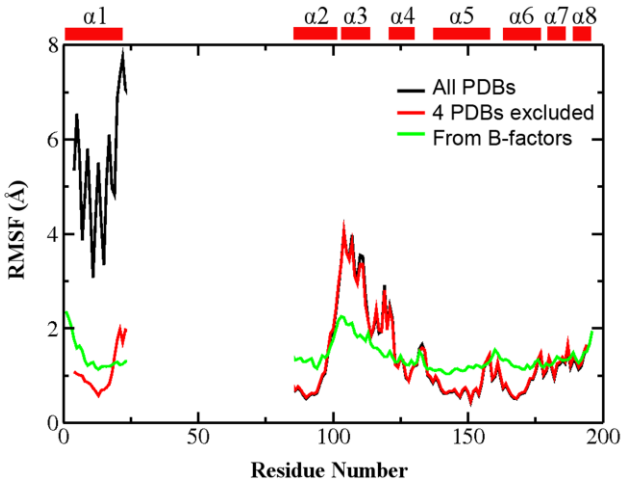


Figure 2.3 RMSF profiles calculated using all 49 Bcl-xL PDB structures (black trace) or 45 Bcl-xL PDB structures excluding 2YQ6, 2YQ7, 3FDL and 4A1U (red trace). The four PDBs excluded form domain-swapped dimers with distinct α1 configurations (see Figure 2.9). The green trace is converted from the B-factors of one representative PDB structure (1R2G). Only Cα atoms of residue 4-23 and 85-194 were used to calculate RMSF since these residues are present in all PDB entries. All structures were aligned using Cα atoms in the core region (residues 85-98, 123-127, 140-156 and 162-175) prior to RMSF calculation.

Clustering analysis was performed to analyze the conformational distribution of the BH3-only protein binding interface of Bcl-xL. For this, all structures were also first aligned using the highly conserved and minimally varying core region, and then clustered based on mutual C_α RMSD of residues 98-120. The clustering led to many small clusters with a cutoff radius of 2.0 Å, confirming the significant level of conformational heterogeneity at the interface. A total of 7 clusters were obtained when the clustering cutoff was set to 3.0 Å. The results are summarized in Table 2.3 and Figure 2.14. Most unbound structures of Bcl-xL were assigned to the same

cluster (cluster 1), whereas structures in complexes populate the other six clusters. The six PDB structures selected for MD simulations are representative of the four most populated clusters, with 2M04 representing the case with the most dramatic local unfolding of Bcl-xL.

2.3.2 Stability and fluctuation of simulation trajectories

Molecular dynamics simulations were performed in explicit solvent to further investigate the conformational dynamics of Bcl-xL in both unbound and bound states. For this, we focus on two representative structures of unbound Bcl-xL, namely 1R2D (by X-ray crystallography) [127] and 2M03 (by NMR) [109], and four Bcl-xL complexes that involve various BH3-only protein ligands as summarized in Table 2.2. The complexes were selected based on the varying degrees of Bcl-xL structural disorder observed at the binding interface (see Figs. 1 and 2). For example, Bcl-xL undergoes substantial unfolding in both α2 and α3 upon PUMA binding (2M04), but remains similarly structured when bound with BIM (3FDL). The helical structures in α3 appear to be sufficiently distorted both Bcl-xL/BAD (1G5J) and Bcl-xL/BAK (1BXL) complexes to be assigned as "coil" using the standard DSSP classification [160]. Interestingly, α2 of Bcl-xL is extended by an extra turn upon BAK binding (Figure 2.1C). The Bcl-xL/BIM complex (3FDL) was simulated without its domain-swapped partner for the sake of computational efficiency.

All six structures were stable during the simulations that lasted 310 to 730 ns. As shown in Figure 2.4, the core region (see Figure 2.3 caption for definition) deviated no greater than ~2 Å from their corresponding initial conformations in all simulations. Without its domain-swapped partner, packing of α1 is very unstable

during the initial stages of the Bcl-xL/BIM simulation (the 3FDL trace), leading to large fluctuations in the overall RMSD (Figure 2.4A). During the simulation, α1 moves quickly towards the rest of the protein within the first 20 ns, and undergo additional major conformational transitions around 170 ns and 250 ns (see Figure 2.12 and Figure 2.13). However, the final configuration of α1 in the 3FDL simulation remains quite different from the typical configuration observed in Bcl-xL monomer structures (see Figure 2.17), which likely reflects limitations in conformational sampling with the 450 ns production run in explicit solvent. Moderate overall RMSD values of the 2M04 and 1R2D trajectories mainly arise from the unfolded BH3-only protein binding interface and fluctuations in the packing of α1, respectively. The structures of the bound BH3-only peptides were also very stable, except that PUMA helix becomes unfolded at the C-terminus (see Figure 2.17 and Figure 2.18). Interestingly, PUMA C-terminal spontaneous unfolding has also been implicated when bound to another Bcl-2 protein Mcl-1 and was shown to facilitates its dissociation from Mcl-1 [161].

Once establishing the stabilities of all PDB structures during the simulations, we quantify the conformational dynamics of Bcl-xL by performing RMSF analysis. For this, all conformations in each trajectory were first aligned using C_α atoms in the core region. The resulting RMSF profiles, shown in Figure 2.5, are highly similar to those extracted from existing PDB strutures (Figure 2.3). The implication is that variations among PDB structures indeed reflect the intrinsic conformational dynamics of Bcl-xL. The BH3-only protein binding interface (e.g., residues 98-120) is the most flexible region in all simulations, except that the α1 segment is artificially

more dynamic in the 3FDL trajectory due to the absence of the domain-swapped partner. We note that the α3 region shows significantly larger flluctations compared with other helices even for structures with well-formed α3 helix (e.g., 2M03, blue trace). Interestingly, binding of various BH3-only peptides do not appear to suppress the interfacial dynamics in general. Instead, the interface tends to become even more dynamic upon binding, e.g., comparing 1BXL (black trace) and 1G5J (red trace) versus 1R2D (green trace). Enhancement in interfacial dynamics is particularly dramatic in the case of Bcl-xL/PUMA complex due to complete local unfolding (cyan trace).

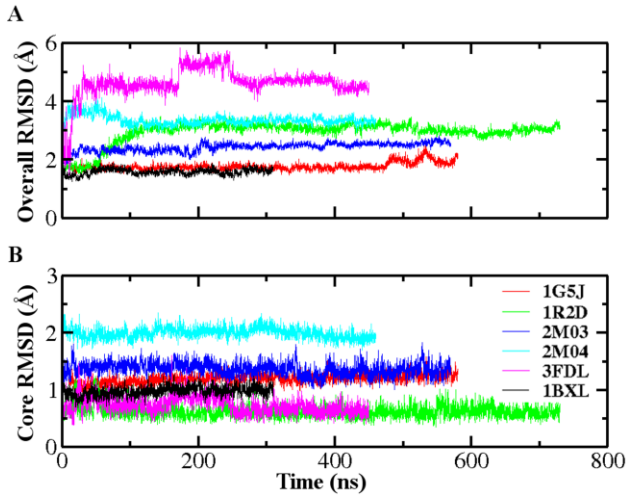


Figure 2.4 Evolution of Bcl-xL backbone heavy atom RMSD for A) the whole protein and B) core region during six simulations initiated from various PDB structures. See Figure 2.3 caption for the definition of the core region.

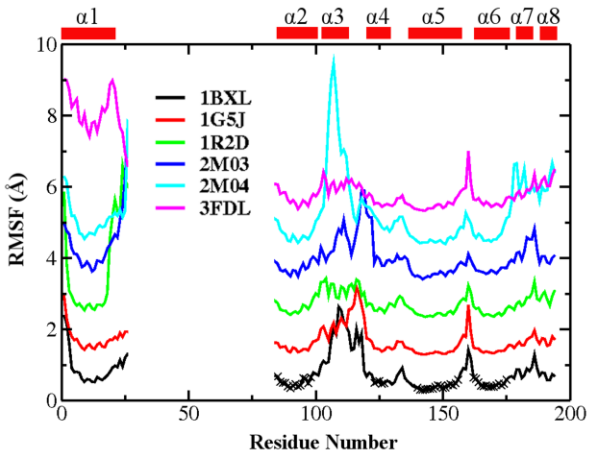


Figure 2.5 RMSF of Bcl-xL derived from all six simulations. Only Cα residues 1-26 and 85-194 are shown since this is the common region among all simulation systems. The core region is marked with ×. RMSF profiles are consecutively shifted by 1.0 Å along the y-axis for clarity.

2.3.3 Conformational dyanamics of the BH3-only Protein Binding Interface

To further examine the details of conformational fluctuation at the BH3-only protein binding interface of Bcl-xL, we first focus on the stability and fluctuation at the secondary structure level. The results, summarized in Figure 2.6 and Figure 2.19, reveal that the α3 segment can undergo spontaneous helix-coil changes in both bound and unbound states. For example, for the Bcl-xL/PUMA complex, α3 is completely unfolded in the original PDB structure (2M04), but starts to sample short helical structures after about 300 ns of simulations. The average residue helicity in the α3 segment calculated from the second half of the 2M04 simulation trajectory reach ~0.3 (Figure 2.7). Similar observation can be also made in simulations initiated from the 1G5J and 1BXL structures, except that the reformation of short helixes in the α3 segment occurred much earlier during the simulations. The neighboring region of α3, including both the N-terminus of α4 and C-terminus of α2, is dynamic as well. For example, α4 N-terminus of the unbound Bcl-xL became

unfolded after about 350 ns in the trajectory initiated from 2M03, but was extended after 75 ns in the trajectory initiated from 2M04. The averaged residue helicity profiles calculated from these trajectories, shown in Figure 2.7 and Figure 2.20, again illustrate the intrinsic dynamics and conformational heterogeneity of the BH3-only protein binding interface of Bcl-xL.

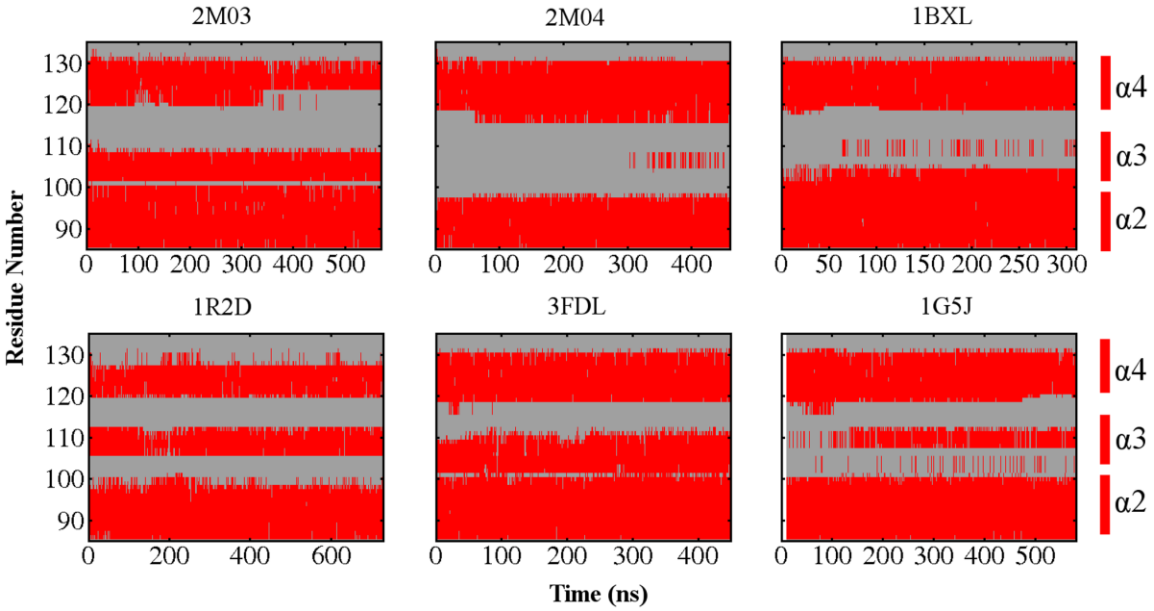


Figure 2.6 Evolution of secondary structures near the BH3-only protein binding interface of Bcl-xL during six sets of simulations. The coloring scheme is the same as in Figure 2. See Figure 2.19 for the results for all Bcl-xL residues.

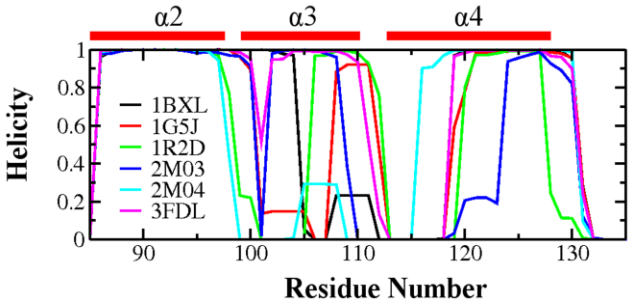


Figure 2.7 Average residue helicity profiles of Bcl-xL calculated from the second half of the simulation trajectories. Only residues 85-135 were shown here for clarity. See Figure 2.20 for the full residue helicity profiles.

Conformations of the BH3-only protein binding interface sampled during the atomistic simulations can be further visualized using the principal component analysis (PCA). For this, all snapshots from both the PDB set and six simulation trajectories were first aligned using the core region (see Figure 2.3 caption for definition), and the C_α positions of residues 98-120 were then analyzed. The first two major components derived from the set of 49 PDB structures were used to project all PDB structures and MD snapshots. The results, summarized in Figure 2.8, show that PDB structures assigned to different clusters nicely segregate on the 2D projection as expected. All six atomistic simulations appear to mainly sample local conformational spaces centered around their corresponding initial conformations. The only exception is the 2M04 trajectory. With a completely unfolded α3, the interfacial structure of 2M04 quickly deviates from the initial conformation (marked by the blue ⊗ in Figure 2.8B), and mainly sample a moderately large space projected to overlap with clusters 3 and 4 (orange and purple x's). We note that the 2MO4 ensemble is quite heterogeneous but most of its members (particularly model 1 used to initiate the simulation) do not overlap with the final conformational space sampled (Figure 2.21). The limited conformational sampling could suggest that the interfacial structures of Bcl-xL is restricted to different conformational subspace depending on the binding partner, but is also likely a direct consequence of the finite simulation times in explicit solvent. Particularly for unbound Bcl-xL (2M03 and 1R2D trajectories), spontaneous unwinding and reforming of interfacial helices and changes in interfacial tertiary structures likely occur at timescales beyond those accessed by the current

simulations. More advanced sampling technique might be required to reliably probe the conformational space accessible to Bcl-xL in bound and unbound states [162-164].

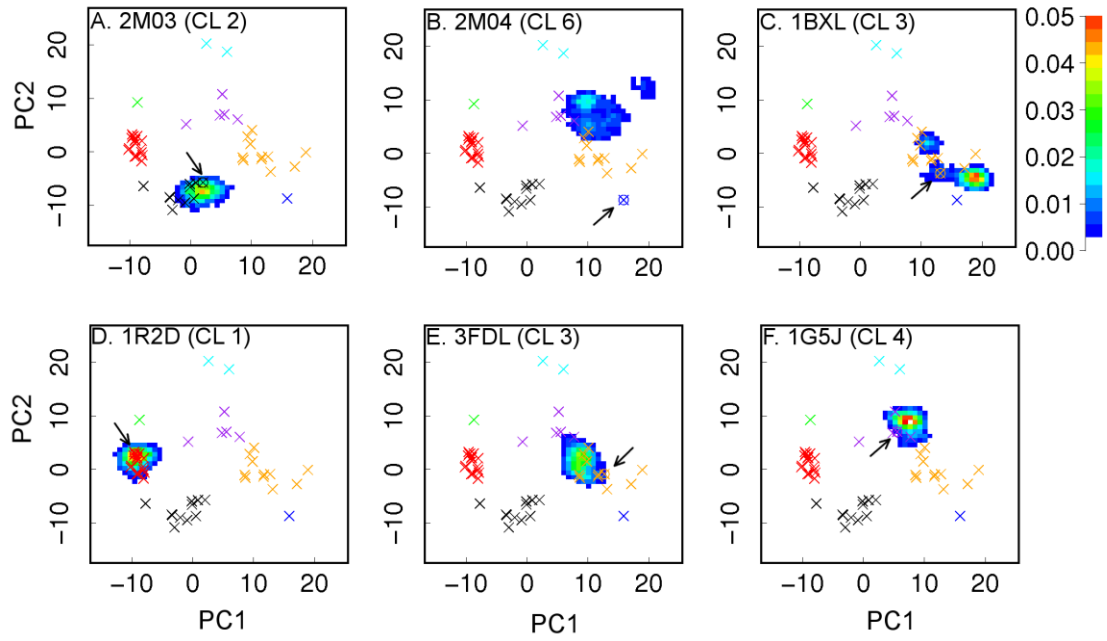


Figure 2.8 Projection of six independent simulations trajectories on the first two major components derived from the set of 49 PDB structures. The projected locations of all PDB structures are marked using ×, except that the starting structure of each simulation is marked with ⊗ and pointed by arrows. Clusters 1 through 7 are colored in red, black, orange, purple, cyan, blue, and green, respectively. The cluster numbers of the PDB structures used to initiate the simulations are also indicated.

2.4 Conclusions

It has been increasingly recognized that regulated unfolding of proteins is widely involved in cellular signaling and transduction [107]. Bcl-xL, in particular, has been recently shown to undergo dramatic local unfolding upon binding and folding of the intrinsically disordered PUMA protein [109]. There is an important need to understand the molecular principles of how such binding induced folding and

unfolding may be achieved with high reliability and efficiency to be viable in cellular signaling and regulation. Critical analysis of existing PDB structures of Bcl-xL in bound and unbound states has revealed that its BH3-only binding interface is intrinsically more dynamic that the rest of the protein. In particular, the specific π-stacking interaction between PUMA Trp71 and Bcl-xL His113 is clearly required for Bcl-xL interfacial unfolding in the case of PUMA [109], but does not appear necessary for Bcl-xL partial unfolding itself. Intrinsic dynamics at the BH3-only protein binding interface of Bcl-xL has been further confirmed by atomistic simulations of six representative bound and unbound structures in explicit solvent, even though these simulations are apparently insufficient to sample large-scale spontaneous conformational fluctuations in explicit solvent. Together, the current study supports that the BH3-only protein binding interface of Bcl-xL is highly dynamic and poised to adopt alternative conformations in response to ligand binding, as well as changes in solution conditions and post-translational modifications. Such intrinsic interfacial conformational plasticity is likely the main physical basis of regulated unfolding observed in the Bcl-xL/PUMA interaction. With the ability to interact with numerous pro-apoptotic proteins including p53, Bcl-xL has been targeted by small molecules for cancer therapy [165]. The dynamic nature of its BH3-only binding interface should be a critical consideration in such rational drug design and optimization efforts.

2.5 Supporting Material

Table 2.2 Information about 49 PDB entries

PDB	Main Binding Partner(s)†	Organism	Method	Mutations*
2ME8	in p53-bound state	Homo sapiens	Solution NMR	
2MEJ	p53	Homo sapiens	Solution NMR	
4A1U	α-β-foldamer 2C	Homo sapiens	X-ray Diffraction	
3PL7	BAX BH3 peptide	Homo sapiens	X-ray Diffraction	
2LP8	BAK BH3 peptide	Homo sapiens	Solution NMR	
1G5J	BAD BH3 peptide	Homo sapiens	Solution NMR	
1BXL	BAK BH3 peptide	Escherichia coli	Solution NMR	
1PQ1	BIM BH3 peptide	Mus musculus	X-ray Diffraction	*A168S, E193D*
2YQ7	BIM BH3 peptide	Homo sapiens	X-ray Diffraction	
3FDL	BIM BH3 peptide	Homo sapiens	X-ray Diffraction	
2PON	Beclin-1 BH3 peptide	Homo sapiens	Solution NMR	
2YQ6	BIM BH3 peptides	Homo sapiens	X-ray Diffraction	
2M04	PUMA BH3 peptide	Homo sapiens	Solution NMR	
2BZW	full-length BAD	Mus musculus	X-ray Diffraction	*A168S, E193D*
1YSN	43B	Homo sapiens	Solution NMR	
1YSI	N3B	Homo sapiens	Solution NMR	
1YSG	4FC and TN1	Homo sapiens	Solution NMR	
3QKD	HI0	Homo sapiens	X-ray Diffraction	
4EHR	0Q5	Homo sapiens	X-ray Diffraction	
2O2N	LIW	Homo sapiens	Solution NMR	
2O2M	LI0	Homo sapiens	Solution NMR	
2O1Y	43B	Homo sapiens	Solution NMR	
2YXJ	N3C	Homo sapiens	X-ray Diffraction	
3SPF	B50	Homo sapiens	X-ray Diffraction	
3SP7	03B	Homo sapiens	X-ray Diffraction	
3ZLO	X8U	Homo sapiens	X-ray Diffraction	
3ZLN	H0Y	Homo sapiens	X-ray Diffraction	
3ZK6	H1I	Homo sapiens	X-ray Diffraction	
2ME9	unbound	Homo sapiens	Solution NMR	
1LXL	unbound	Homo sapiens	Solution NMR	
1MAZ	unbound	Homo sapiens	X-ray Diffraction	
3CVA	unbound	Homo sapiens	X-ray Diffraction	W137A
1PQ0	unbound	Mus musculus	X-ray Diffraction	*A168S, E193D*
3ILB	unbound	Mus musculus	X-ray Diffraction	R139A, *A168S, E193D*
3ILC	unbound	Mus musculus	X-ray Diffraction	Y101A, *A168S, E193D*
3IIH	unbound	Mus musculus	X-ray Diffraction	*A168S, E193D*
2LPC	unbound	Homo sapiens	Solution NMR	
1AF3	unbound	Rattus norvegicus	X-ray Diffraction	*A168S, E193D*
1R2I	unbound	Homo sapiens	X-ray Diffraction	F146L
1R2H	unbound	Homo sapiens	X-ray Diffraction	A142L
1R2G	unbound	Homo sapiens	X-ray Diffraction	F97W
1R2E	unbound	Homo sapiens	X-ray Diffraction	E92L
1R2D	unbound	Homo sapiens	X-ray Diffraction	

2M03	unbound	Homo sapiens	Solution NMR	
3IIG	unbound	Mus musculus	X-ray Diffraction	F105A, *A168S, E193D*
3IHF	unbound	Mus musculus	X-ray Diffraction	R139A, *A168S, E193D*
3IHE	unbound	Mus musculus	X-ray Diffraction	F105A, *A168S, E193D*
3IHD	unbound	Mus musculus	X-ray Diffraction	Y101A, *A168S, E193D*
3IHC	unbound	Mus musculus	X-ray Diffraction	*A168S, E193D*

*Mouse Bcl-xL is different from human Bcl-xL that A168 and E193 are S168 and D193, respectively. Here, all mutations are shown with respect to 2M04 amino acid sequence.

† Full name of small molecule ligand are as follows:

43B: 3-nitro-n-{4-[2-(2-phenylethyl)-1,3-benzothiazol-5-yl]benzoyl}-4-{[2-(phenylsulfanyl)ethyl]amino}benzenesulfonamide

N3B: n-[(4'-fluoro-1,1'-biphenyl-4-yl)carbonyl]- 3-nitro-4-{[2-(phenylsulfanyl)ethyl]amino}benzenesulfonamide

4FC: 4'-fluoro-1,1'-biphenyl-4-carboxylic acid

TN1: 5,6,7,8-tetrahydronaphthalen-1-ol

HI0: (R)-N-(7-(4-((4'-chlorobiphenyl-2-yl)methyl)piperazin- 1-yl)quinazolin-4-yl)-4-(4-(dimethylamino)- 1-(phenylthio)butan-2-ylamino)-3-nitrobenzenesulfonamide

0Q5: 4-[5-butyl-3-(hydroxymethyl)-1-phenyl-1H- pyrazol-4-yl]-3-(3,4-dihydroisoquinolin-2(1H)- ylcarbonyl)-N-{[2-(trimethylsilyl)ethyl]sulfonyl}benzamide

LIW: 4-[4-(biphenyl-2-ylmethyl)piperazin-1-yl]- n-[(4-{[1,1-dimethyl-2-(phenylthio)ethyl]amino}- 3-nitrophenyl)sulfonyl]benzamide

LI0: 4-(4-benzyl-4-methoxypiperidin-1-yl)-n-[(4- {[1,1-dimethyl-2-(phenylthio)ethyl]amino}- 3-nitrophenyl)sulfonyl]benzamide

N3C: 4-{4-[(4'-chlorobiphenyl-2-yl)methyl]piperazin- 1-yl}-n-{[4-({(1r)-3-(dimethylamino)-1-[(phenylthio)methyl]propyl}amino)- 3-nitrophenyl]sulfonyl}benzamide

B50: 4-(4-chlorophenyl)-1-[(3S)-3,4-dihydroxybutyl]- N-[3-(4-methylpiperazin-1-yl)propyl]-3-phenyl- 1H-pyrrole-2-carboxamide

03B: 5-(4-chlorophenyl)-4-{3-[4-(4-{[(4-{[(2R)- 4-(dimethylamino)-1-(phenylsulfanyl)butan- 2-yl]amino}-3-nitrophenyl)sulfonyl]amino}phenyl)piperazin- 1-yl]phenyl}-1,2-dimethyl-1H-pyrrole-3-carboxylic acid

X8U: 2-[(8E)-8-(1,3-benzothiazol-2-ylhydrazinylidene)- 6,7-dihydro-5H-naphthalen-2-yl]-5-(4-phenylbutyl)- 1,3-thiazole-4-carboxylic acid

H0Y: 6-[(8E)-8-(1,3-benzothiazol-2-ylhydrazinylidene)-6,7-dihydro-5H-naphthalen-2-yl]pyridine-2- carboxylic acid

H1I: N-(3-(5-(1-(2-(benzo[d]thiazol-2-yl)hydrazono)ethyl)furan-2-yl)phenylsulfonyl)-6-phenylhexanamide

Table 2.3 Clustering results of the 49 PDB structures ordered by cluster size. The structures of all clusters are shown in Figure 2.11. PDB structures selected for simulations are highlighted in red (also see Table 2.2).

PDB	Main Binding Partner(s)†	Cluster Number	PDB	Main Binding Partner(s)†	Cluster Number
1AF3	unbound	1	2O1Y	43B	2
1MAZ	unbound	1	3ZLN	H0Y	2
1PQ0	unbound	1	3ZLO	X8U	2
1R2D	unbound	1	3ZK6	H1I	2
1R2E	unbound	1	3QKD	HI0	3
1R2G	unbound	1	2YXJ	N3C	3
1R2H	unbound	1	3SPF	B50	3
1R2I	unbound	1	4EHR	0Q5	3
3CVA	unbound	1	2YQ6	BIM BH3 peptides	3
3IHC	unbound	1	2YQ7	BIM BH3 peptide	3
3IHD	unbound	1	3FDL	BIM BH3 peptide	3
3IHE	unbound	1	3PL7	BAX BH3 peptide	3
3IHF	unbound	1	1BXL	BAK BH3 peptide	3
3IIG	unbound	1	2PON	Beclin-1 BH3 peptide	3
3IIH	unbound	1	4A1U	α-β-foldamer 2C	3
3ILB	unbound	1	2LP8	BAK BH3 peptide	4
3ILC	unbound	1	3SP7	03B	4
1LXL	unbound	2	1G5J	BAD BH3 peptide	4
2LPC	unbound	2	1PQ1	BIM BH3 peptide	4
2M03	unbound	2	2BZW	full-length BAD	4
1YSG	4FC and TN1	2	2ME8	in p53-bound state	5
1YSI	N3B	2	2MEJ	p53	5
1YSN	43B	2	2M04	PUMA BH3 peptide	6
2O2M	LI0	2	2ME9	unbound	7
2O2N	LIW	2			

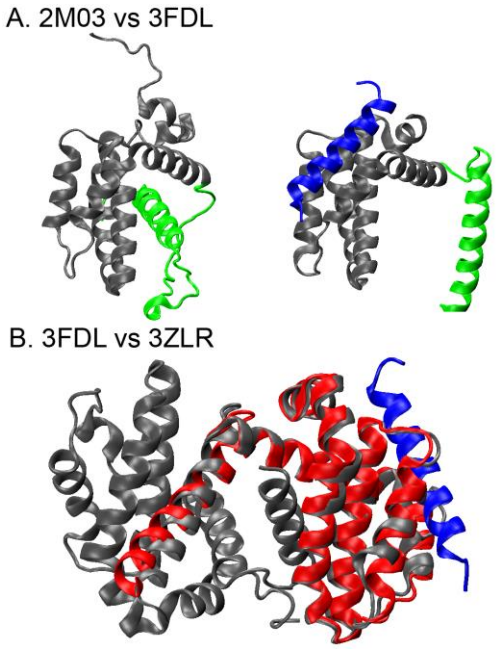


Figure 2.9 A) Packings of α1 in Bcl-xL monomer (2M03 as an example) and domain-swapped dimers (3FDL as an example). α1 is colored in green. B) An illustration of the domain-swapped structures of Bcl-xL. For 3FDL, Bcl-xL is colored in red and BIM in blue. The domain-swapped Bcl-xL dimers in 3ZLR are colored in gray.

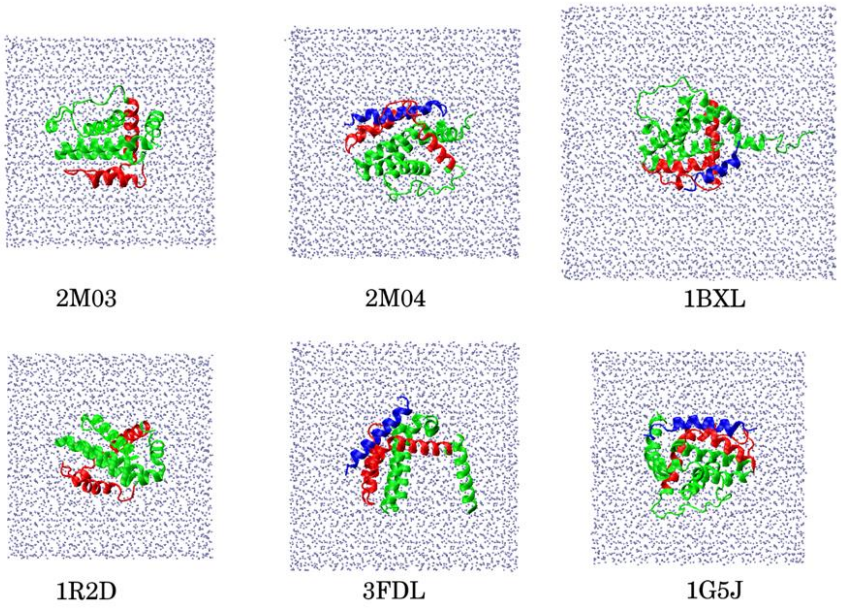


Figure 2.10 Solvated simulation boxes of all six systems. Water molecules are shown in iceblue. Bcl-xL is shown in Cartoon representation, colored in green except for α2, α3 and α4, which are colored in red. The BH3-only protein ligands are shown in blue.

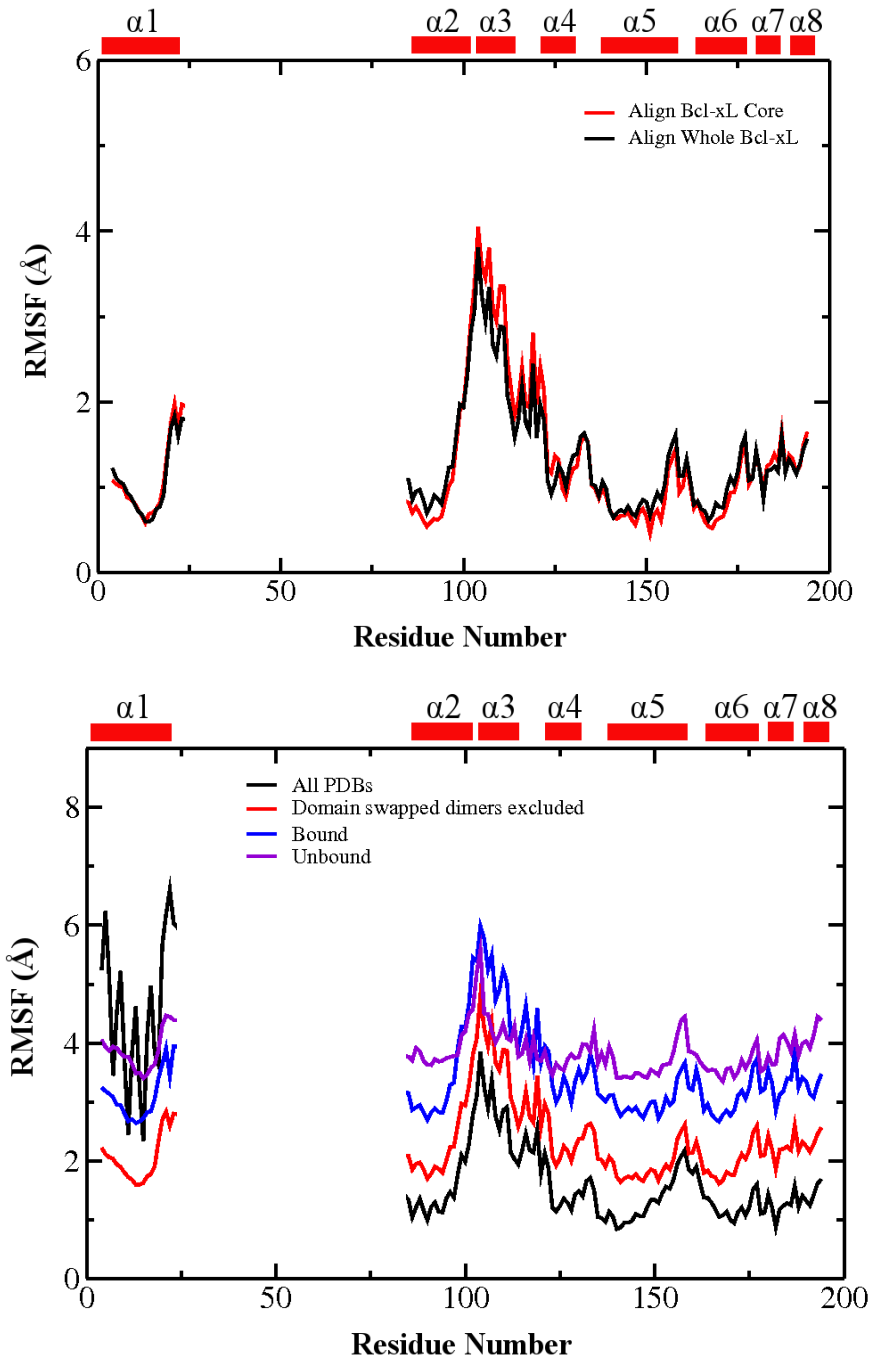


Figure 2.11 Top: RMSF profiles calculated with PDB structures aligned using either the core region (black trace) or the whole protein (core trace). Only the 45 monomeric Bcl-xL PDB structures were included. Bottom: RMSF profiles calculated from 24 bound (blue trace) and 21 unbound structures of Bcl-xL (violet trace). RMSF profiles are consecutively shifted by 1.0 Å along the y-axis for clarity.

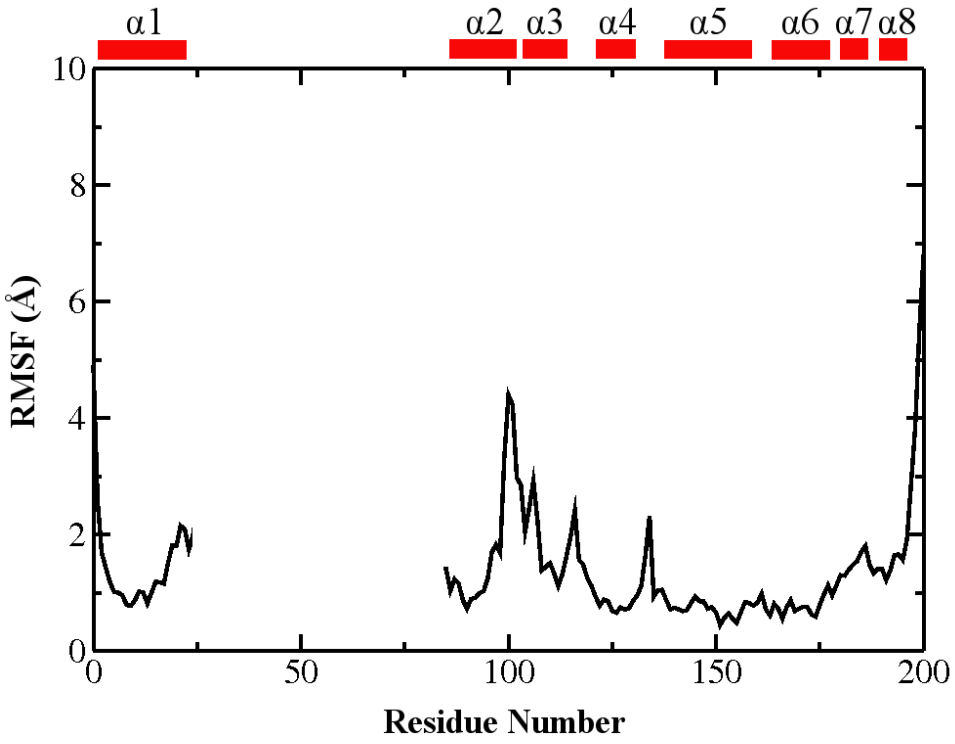


Figure 2.12 RMSF profiles calculated from 20 Bcl-xL PDB structures in 2M04. All structures were aligned using Cα atoms in the core region (residues 85-98, 123-127, 140-156 and 162-175) prior to RMSF calculation.

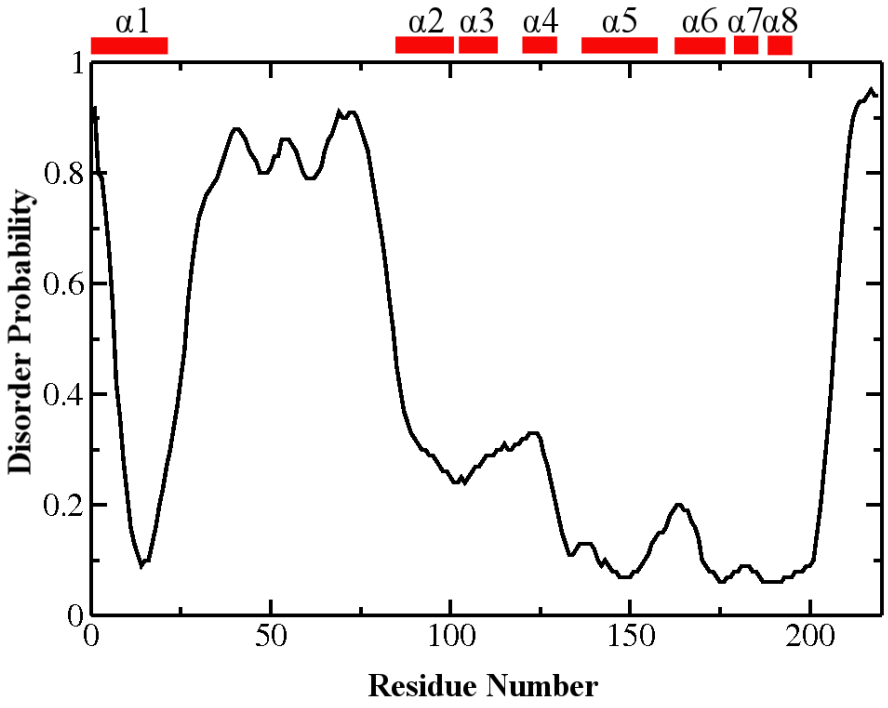


Figure 2.13 Intrinsic disorder propensity of Bcl-xL predicted by PrDOS.

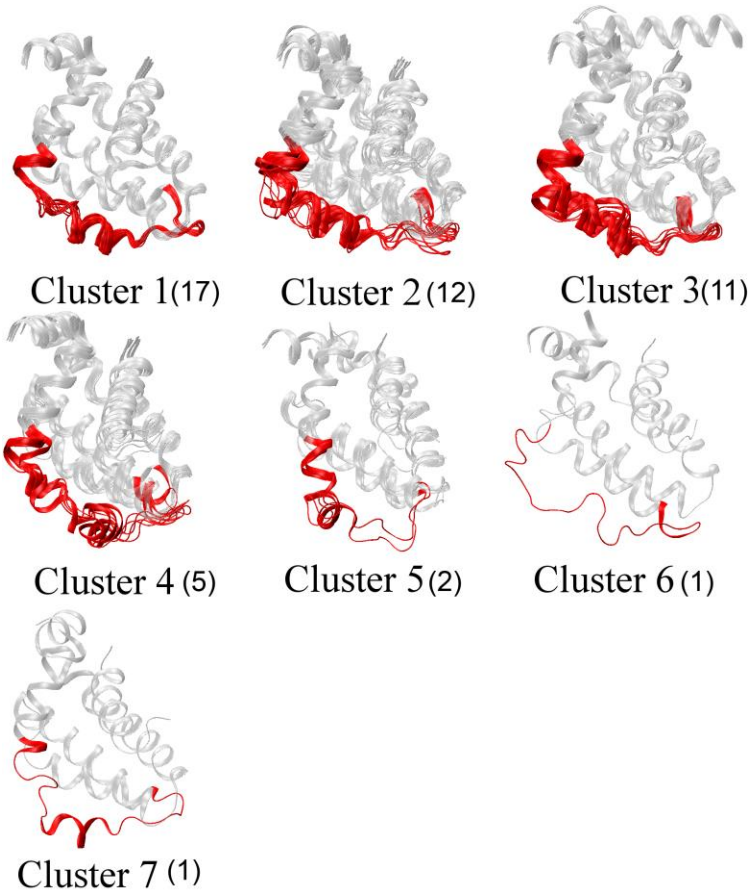


Figure 2.14 Clustering of the 49 PDB structures. All structures are aligned using the core region as defined in the main text before clustered based on the Cα RMSD of residues 98-120 (colored in red). The bound ligand (if present) is not shown for clarity. The cluster memberships of all PDBs are listed in Table 2.3.

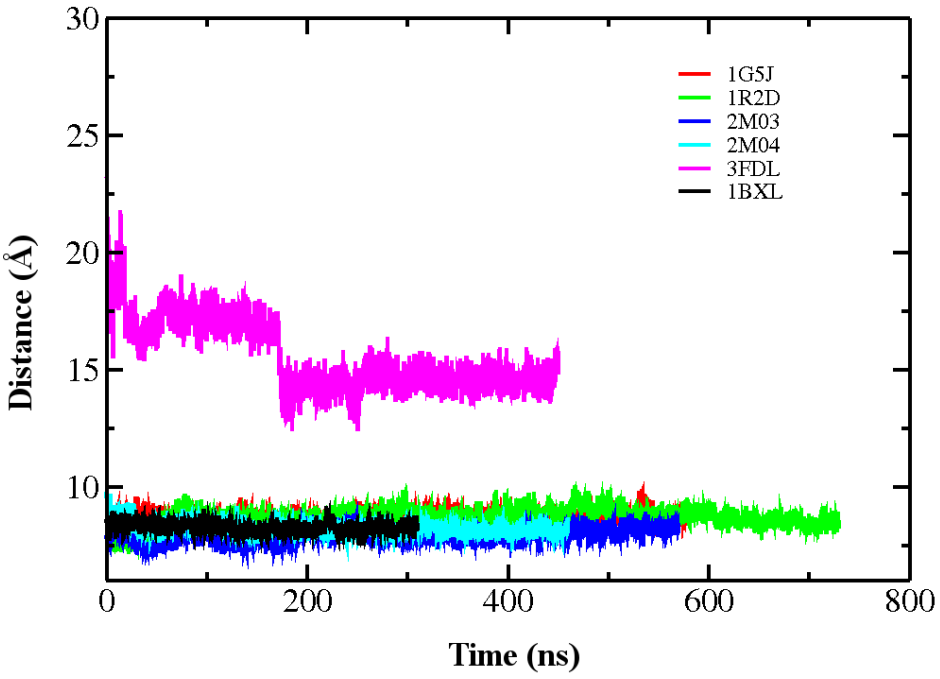


Figure 2.15 Time evolution of the center of mass (COM) distance between α1 and the core region of Bcl-xL during the six independent simulations.

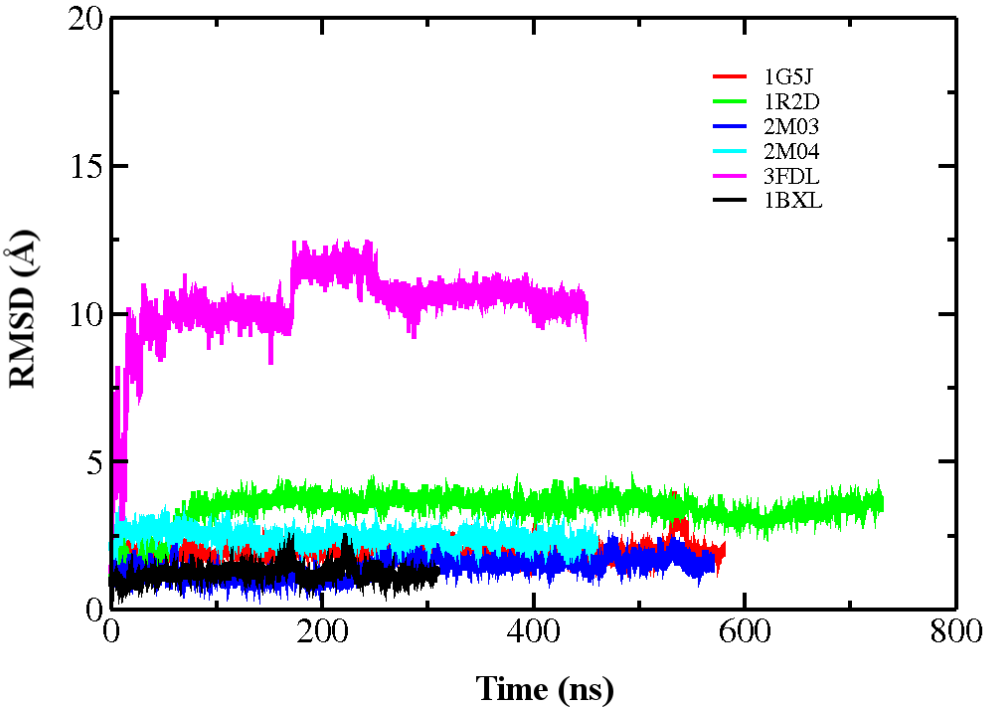


Figure 2.16 Time evolution of the backbone RMSD of Bcl-xL α1 during the simulations. All snapshots were aligned to the initial conformations using the core region.

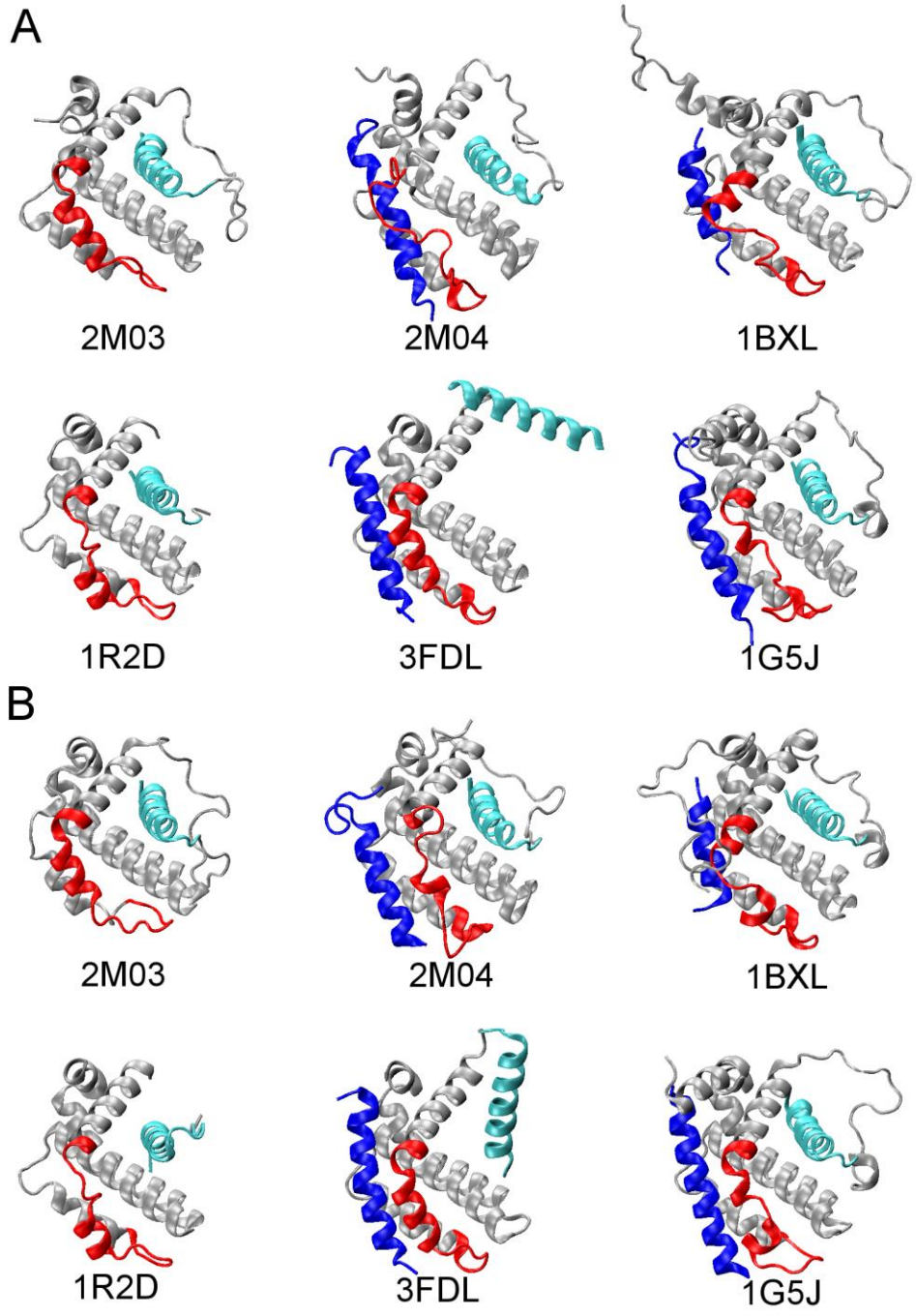


Figure 2.17 Starting (A) and final (B) conformations of all six simulations. α1 is colored in cyan, residues 98-120 in red, and bound peptide in blue.

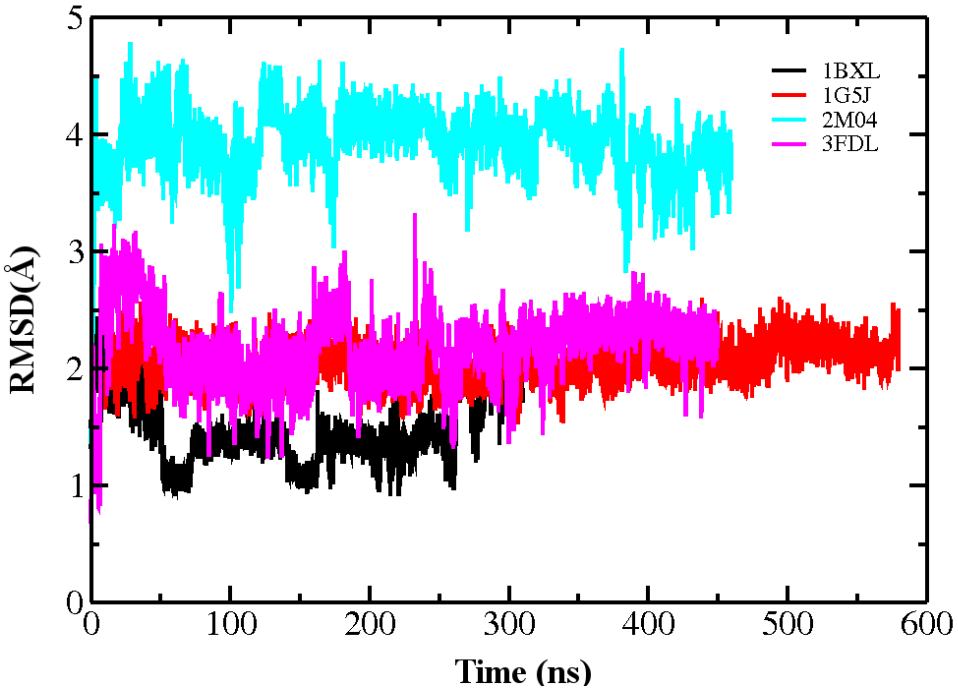


Figure 2.18 Time evolution of the backbone RMSD of BH3-only peptide ligands during the simulations. All snapshots were aligned to the initial conformations using the Bcl-xl core region (as defined in the main text).

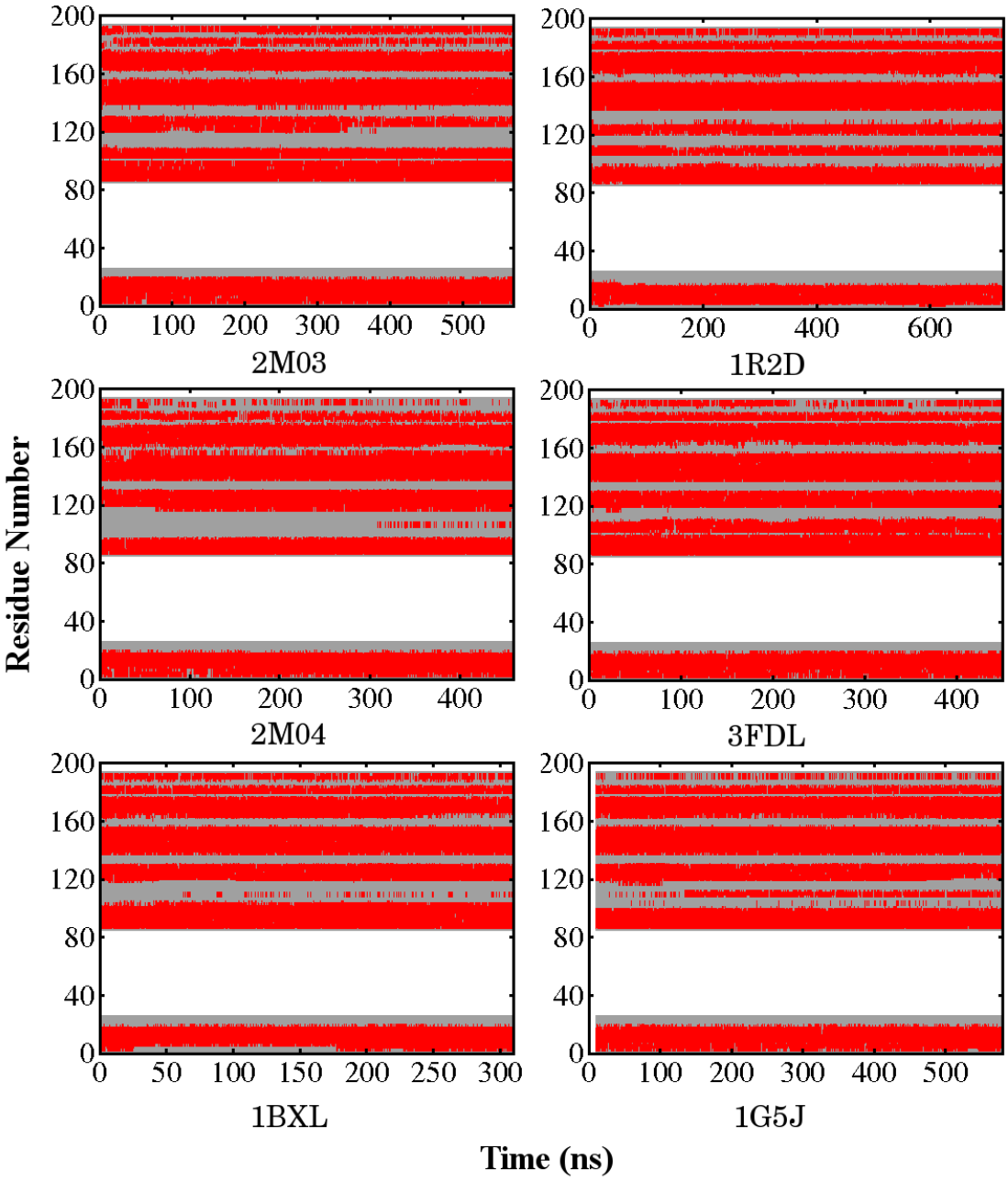


Figure 2.19 Time evolution of Bcl-xL secondary structure for each simulation trajectory. Red indicates helical, gray indicates disordered, and white corresponds to deleted regions.

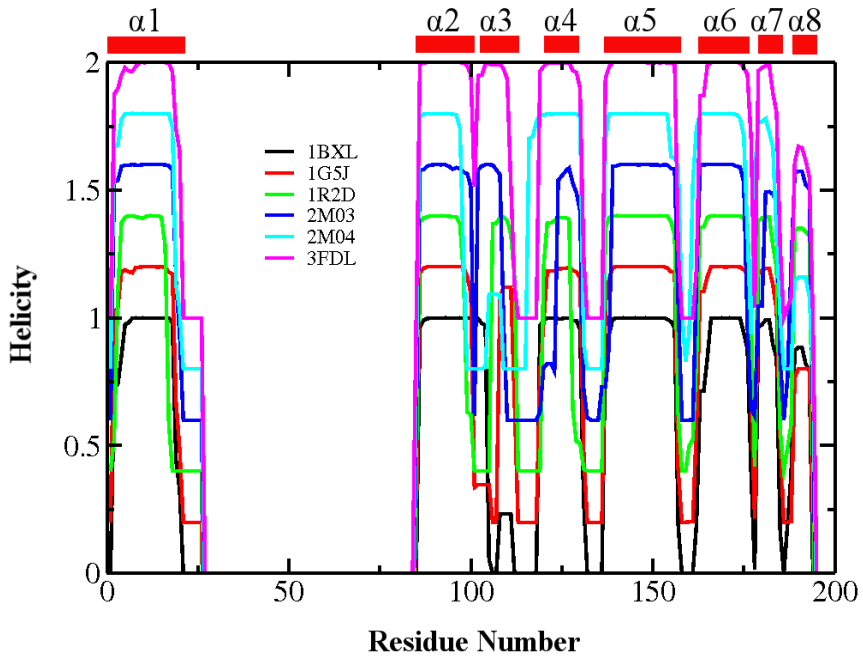


Figure 2.20 Average residue helicity of Bcl-xL derived from the second half of the simulation trajectories. Incremental vertical offset of 0.2 was added for clarity.

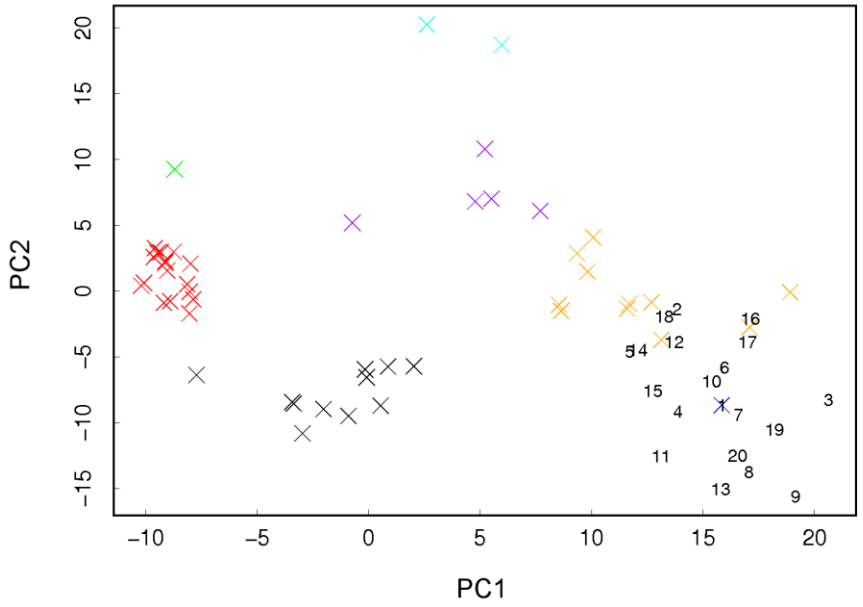


Figure 2.21 Projection of 20 PDB structures in 2M04 on the first two major components derived from the set of 49 PDB structures. The projected locations of the 49 PDB structures are marked using ×, while the 20 PDB structures in 2M04 are shown with their respective model numbers. Clusters 1 through 7 are colored in red, black, orange, purple, cyan, blue, and green, respectively.

CHAPTER 3

ENHANCED SAMPLING OF INTRINSIC STRUCTURAL HETEROGENEITY OF THE BH3-ONLY PROTEIN BINDING INTERFACE OF BCL-XL*

Anti-apoptotic Bcl-xL plays central roles in regulating programed cell death. Partial unfolding of Bcl-xL has been observed at the interface upon specific binding to the pro-apoptotic BH3-only protein PUMA, which in turn disrupts the interaction of Bcl-xL with tumor suppressor p53 and promotes apoptosis. Previous analysis of existing Bcl-xL structures and atomistic molecular dynamics (MD) simulations have suggested that substantial intrinsic structure heterogeneity exists at the BH3-only protein binding interface of Bcl-xL to facilitate its conformational transitions upon binding. In this study, enhanced sampling is applied to further characterize the interfacial conformations of unbound Bcl-xL in explicit solvent. Extensive replica exchange with solute tempering (REST) simulations, with a total accumulated time of 16 μs, were able to cover much wider conformational spaces for the interfacial region of Bcl-xL. The resulting structural ensembles are much better converged, with local and long-range structural features that are highly consistent with existing NMR data. These simulations further demonstrate that the BH3-only protein binding interface of Bcl-xL is intrinsically disordered and samples many rapidly interconverting conformations. Intriguingly, all previously observed conformers are well represented in the unbound structure ensemble. Such intrinsic structural heterogeneity and flexibility may be critical for Bcl-xL to undergo partial unfolding

* Liu X, Jia Z and Chen J (2017). "Enhanced Sampling of Intrinsic Structural Heterogeneity of the BH3-Only Protein Binding Interface of Bcl-xL." Journal of Physical Chemistry B 121(39): 9160-9168

induced by PUMA binding, and likely provide a robust basis that allows Bcl-xL to respond sensitively to binding of various ligands in cellular signaling and regulation.

3.1 Introduction

Intrinsically disordered proteins (IDPs) in their unbound state do not form well-defined three-dimensional structures under physiological conditions[2, 3, 166], in contrast to the conventional protein structure-function paradigm[1]. They are highly prevalent in biology[167-169] and play critical roles in cellular signaling and regulation[13, 17, 21, 170, 171]. Mutations of IDPs[172-174] and/or changes in their protein levels[170, 175, 176] have been implicated in numerous human diseases. There is thus an increasing interest in understanding the physical properties of IDPs and how these properties contribute to versatile functions. In particular, inherent structural fluctuations of IDPs in their unbound states are likely key to understanding how IDPs could respond rapidly and sensitively to various stimuli in cellular processes[177, 178].

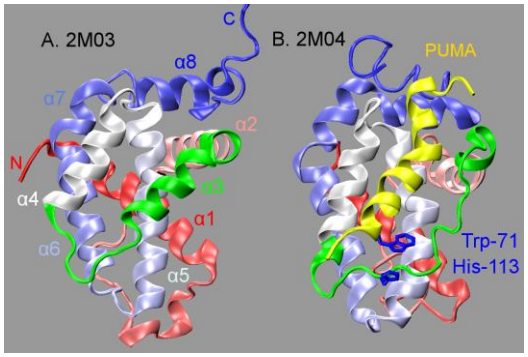


Figure 3.1 PDB structures of unbound (A) and PUMA-bound (B) Bcl-xL. Bcl-xL is shown in cartoon representation with its color changing from red (at N-terminus) to blue (at C-terminus). The BH3-only protein binding interface (residues 98-120) is highlighted in green, and the bound PUMA is shown in yellow.

Coupled binding and folding of proteins is frequently observed in IDP interactions, during which disordered IDPs fold into stable three-dimensional structures upon specific binding[100, 161, 179]. It has been recently recognized that regulated unfolding of proteins is also often involved in cell signaling[107]. A particularly interesting example involves Bcl-xL, a pro-survival Bcl-2 family protein that regulates programed cell death[180]. Bcl-xL could become partially unfolded at the BH3-only protein binding interface upon specific binding to PUMA, a pro-apoptotic BH3-only Bcl-2 family protein. Note that PUMA contains intrinsically disordered regions and its BH3 domain folds into a helix upon specific binding to Bcl-xL (see Figure 3.1). Partial unfolding of Bcl-xL disrupts its interaction with tumor suppressor p53, which in turn abolishes p53 inhibition of Bcl-xL pro-survival functions and activates the apoptotic cascade[118, 119, 181]. To date, the molecular mechanisms of PUMA-induced partial unfolding of Bcl-xL have not been fully understood. Trp-71 in PUMA has been suggested to play critical roles in driving Bcl-xL local unfolding, which interacts with Bcl-xL His-113 through π-stacking interactions (Figure 3.1)[181]. However, the π-stacking interaction itself appears to contribute little to the binding of Bcl-xL with PUMA, since $PUMA^{W71A}$ mutant associates with wild type Bcl-xL with similar affinities[181]. In addition, the BH3 domain of another BH3-only Bcl-2 family protein, BAD[182], also contains a Trp that could form similar Trp-His contact with Bcl-xL, but its binding does not induce Bcl-xL unfolding[181].

Our previous analysis of existing experimental structures of Bcl-xL in both unbound and bound states and atomistic molecular dynamics (MD) simulations have

suggested that substantial intrinsic structural heterogeneity exists at the BH3-only protein binding interface of Bcl-xL[94], which could provide a robust molecular basis to support Bcl-xL conformational transitions in response to ligand binding. However, atomistic simulations of the heterogeneous structural ensemble of the disordered interface of Bcl-xL is extremely challenging. Conventional MD simulations of ~300-700 ns at the room temperature did not yield converged conformational ensembles[94]. As illustrated in Figure 3.2 C and D, these simulations mainly sample local conformational space near their corresponding initial structures, and conformational spaces visited by different MD simulations do not overlap with each other. In this work, we utilize an advanced sampling technique known as replica exchange with solute tempering (REST)[183, 184] to calculate better converged structural ensembles of Bcl-xL in the unbound state. REST is a Hamiltonian replica exchange method, where the solute and solute-solvent energies are scaled by λ and $\sqrt{\lambda}$, respectively. For the condition λ=1, the system temperature is the one of interest (T); whereas if λ<1, the effective temperature of the solute is increased to T/λ, while the solvent temperature remains at T. As designed, the exchange acceptance ratio is determined by solute energy and solute-solvent energies alone, but not the solvent energy (which is the major component of the total energy). Therefore, much fewer replicas could be used to achieve sufficient energy overlap between neighboring replicas compared with the regular temperature replica exchange[78, 185]. Furthermore, REST allows us to focus tempering on any molecular region of interest, such as the binding interface of Bcl-xL. These properties together make REST particularly suitable for enhanced sampling of larger

systems in explicit solvent with significantly reduced computational cost[183, 184, 186-189]. In the present study, the improved sampling capability of REST has allowed the generation of much better converged structure ensembles of the BH3-only protein binding interface of Bcl-xL. The calculated ensembles have been validated by direct comparison with published NMR data, shedding light on how Bcl-xL could respond sensitively and rapidly to various binding ligands, particularly partial unfolding upon PUMA binding.

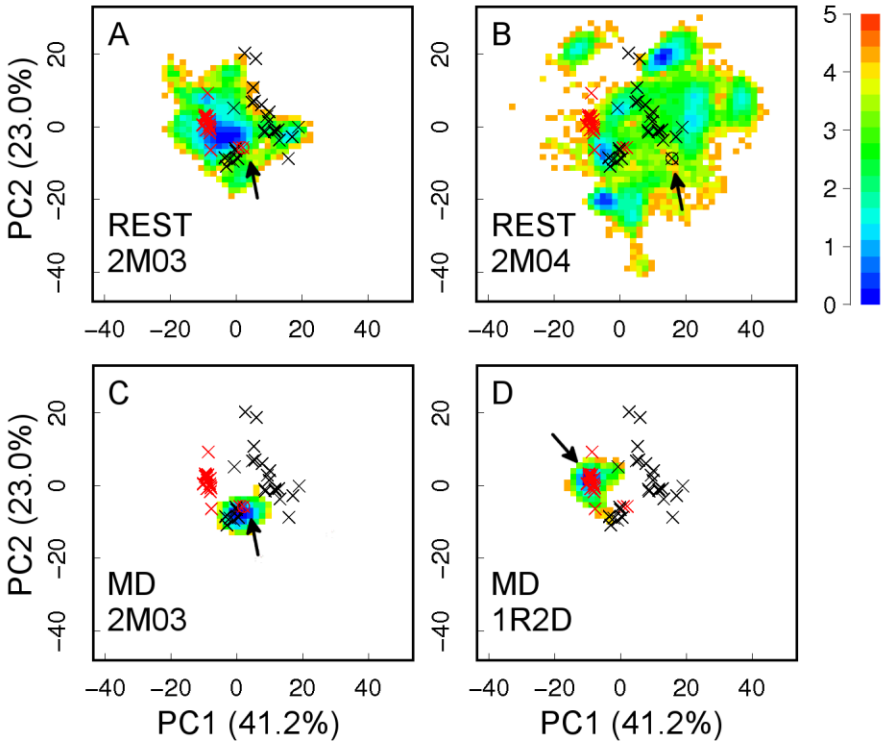


Figure 3.2. Projection of simulation trajectories on the first two principal components (PCs) derived from 49 PDB structures of Bcl-xL. The positions of all PDB structures are marked with ×, with unbound and bound ones in red and black, respectively. The initial structure of each simulation is marked with an arrow. Results from MD simulations are adapted from our previous work[94]. The heat map shows the free energy (in kcal/mol) derived from simulation statistics. Values in the parentheses are percentages of variance associated with each PC.

3.2 Methods

3.2.1 System setup and simulation protocols

Two REST simulations of unbound Bcl-xL have been performed. The initial structures were obtained from RCSB Protein Data Bank (PDB) entries 2M03 (unbound) and 2M04 (PUMA-bound)[181]. The first models from these ensembles were processed by renumbering residues in 2M03 to match that in 2M04. The bound PUMA peptide in 2M04 was removed, and only the common region of Bcl-xL (residues 1-44 and 85-200) were retained. The termini were capped with an acetyl group and N-methyl amide at the N- and C-termini, respectively. Note that the interface region (highlighted in green in Figure 3.1A) exhibits high-helical structure in 2M03, whereas it is almost fully disordered in 2M04 (Figure 3.1B). The significantly different starting structures allow one to better evaluate the level of convergence and other properties obtained from these two independent simulations. The protein was solvated using TIP3P water molecules in a rectangular simulation box constructed with at least 1.0 nm of solvent between protein and the nearest side of the box. 12 Na^+ ions were added to neutralize the system. Details of the system setup are summarized in Table 3.1 and Figure 3.9.

Both simulations (referred to using the initial structures, 2M03 and 2M04, hereafter) were performed using GROMACS-4.6.7[190] patched with PLUMED-2.1.3[188, 191]. To begin with, energy minimization was carried out for both systems using a steepest descent algorithm. Then 50 ps NVT simulation followed by 50 ps NPT simulation was used to equilibrate the system. The equilibrated boxes were then used to initiate two independent REST simulations under NVT conditions. Only the BH3-

only protein binding interface, residues 98-120, was subjected to tempering. The effective temperatures range from 298 K to 500 K, exponentially spaced among 16 replicas. The choice of 500 K as the highest temperature provides accelerated conformational diffusion without requiring a larger number of replicas to cover the temperature span. The corresponding values of the scaling factor λ are 1, 0.966, 0.933, 0.902, 0.871, 0.842, 0.813, 0.785, 0.759, 0.733, 0.708, 0.684, 0.661, 0.639, 0.617, and 0.596, respectively. The frequency of neighbor exchange attempts was every 2 ps, which is commonly used and believed to allow efficient mixing of replicas while ensuring proper equilibration prior to exchange attempts. The length of each REST simulation is 500 ns per replica, and the average exchange acceptance ratio is ~ 50%. In both simulations, weak position restraints with a force constant of 100 kJ/mol/nm^2 were imposed on C_α atoms of the Bcl-xL core region, which includes residues 85-95, 123-127, 140-156, and 162-175. The core region has been shown to display minimal variance among all PDB structures of Bcl-xL[94]. The Amber99sb force field[75] was used. Although it's different from CHARMM36 force field used in our previous MD simulations, both force fields are well parameterized and not expected to significantly affect simulation results. The cutoff distance of van der Waals interactions was set to 1.0 nm, and the nonbonded list was updated every 5 steps. Particle Mesh Ewald (PME) method[155] was used to calculate long-range electrostatic interactions, with a 0.12 nm Fourier spacing. All bonds were constrained using LINCS[192], and the integration time step was 2 fs.

Table 3.1. Summary of system setup

Initial Structure	Water Number	Na^+ Number	Total Atom Number	Simulation Box (nm × nm × nm)
2M03	7361	12	24601	6.59 × 6.54 × 5.69
2M04	8566	12	28216	7.25 × 6.63 × 5.86

3.2.2 NOE, clustering and principal component analyses

Only snapshots sampled at $\lambda = 1$ during the last 200 ns of the simulation were included in NOE analysis. For each frame, NOE-like distances were first calculated based on the r^{-6} summation scheme for cases involving multiple equivalent protons[193, 194]. The averaged NOE distances for simulated ensembles were then calculated based on $\langle r^{-6} \rangle^{-1/6}$ averaging scheme. Experimental NMR restraints of unbound Bcl-xL were obtained from PDB 2M03. Among the 527 total NOE distance restraints, there are 29 NOEs within the interface region (residues 98-120) and 30 NOEs between the interface and the rest of the protein. Only these 59 interface-related NOEs were considered in the assessment of the simulated ensembles. As we compute contact map between residues in the interface region and the rest of the protein, experimental NOE distance restraints from another unbound Bcl-xL structure, PDB 2LPC[141], were also considered for reference.

For principal component analysis (PCA), Bcl-xL structures, either from PDB or simulation trajectories, were first aligned using Cα atoms of the core region (residues 85-98, 123-127, 140-156, and 162-175), and Cα atoms of the interface region (residues 98-120) were then analyzed. The first two principal components (PCs) were derived from 49 PDB structures that include Bcl-xL configurations in both bound and unbound states (see reference[94] for more details). Snapshots from

various simulation trajectories were then projected onto these two PCs. In REST simulations, we extracted conformations every 20 ps from the first condition (λ=1). The free energy surfaces were derived directly from the 2D probability distributions along the first two PCs.

Clustering analysis was performed using the "gmx cluster" tool in GROMACS. All snapshots from two REST simulations were first aligned using backbone heavy atoms of the core region (residues 85-98, 123-127, 140-156, and 162-175), and root-mean-square deviation (RMSD) of backbone heavy atoms at the interface region (residues 98-120) was then used for clustering, with a cutoff value of 3.0 Å.

3.2.3 Conformational entropy

To quantify peptide backbone flexibility and compare sampling efficiency between MD and REST simulations, we calculated the conformational entropy for each residue's backbone dihedral angles φ and ψ. For a given 2D probability distribution of (φ, ψ) with bin size of 8° × 8°, conformational entropy was estimated as

$$S = -k_B \sum p_i ln p_i \quad (3.1)$$

where p_i is the probability of each bin in φ/ψ space, and k_B is the Boltzmann constant.

3.3 Results and Discussion

3.3.1 Enhanced sampling of the BH3-only protein binding interface in REST simulations

Explicit solvent MD simulations initiated from distinct PDB structures at 300 K only managed to sample very limited conformational spaces near the initial conditions[94] (e.g., Figure 3.2C-D). The ability of REST to enhance the sampling

of Bcl-xL interfacial conformation is first evaluated by examining at the secondary structure level. The main structure segment of Bcl-xL that appears to adopt varying conformations is helix α3. During both REST runs, helix α3 underwent numerous helix-coil transitions. As a result, the ensembles sampled at the $\lambda = 1$ condition contain mixtures of various helix and coil conformations regardless of the initial PDB structures (Figure 3.3). The average residue helicity in segment α3 peaks at ~0.75 for simulations initiated from 2M03 and ~0.38 from REST run 2M04 (Figure 3.10). Even though the convergence is still far from ideal, it is much better than the previous regular MD simulations, where helix α3 remained largely folded after up to 710 ns when initiated from PDB structures (e.g., 1R2D) with fully folded helix[94]. The improved sampling ability of REST simulations is also evident by examining the backbone φ/ψ space visited by residues in the interfacial region, most of which sampled broader distributions in REST simulations (Figure 3.11). The peptide conformations appear much less likely to be trapped in local energy minima during REST simulations. As such, residue backbone conformational entropies calculated using Eq. 3.1 are consistently higher from REST simulations (see Figure 3.4). Note that, although residues 104-119 in MD simulation of 2M04, and residues 105-106 in MD simulation of 1BXL show higher conformational entropy than in REST simulations, this is most likely due to the lack of stable interfacial helix reforming in these MD simulations[94].

The quality and convergence of conformational sampling at the tertiary level was evaluated using PCA analysis. As shown in Figure 3.2, REST simulations were able to sample much broader conformational space than any of the previous MD runs.

Furthermore, conformational spaces sampled by both REST runs appear to have significant overlaps, with run 2M04 sampling a substantially broader region that includes what was sampled by run 2M03. Interestingly, all representative conformational sub-states identified from existing PDB structures, including both ligand-bound and ligand-free states, appear to be visited in both REST simulations (Figure 3.2A-B). Nonetheless, we note that substantial differences persist between structural ensembles obtained from two REST runs at both secondary and tertiary levels, suggesting the simulations were not fully converged. This highlights the formidable challenges of sampling complex and heterogeneous conformational space of moderately-sized disordered protein segments, particularly in the context of a bigger protein. It probably requires significantly longer REST simulations, or more efficient advanced sampling techniques to further improve the level of convergence in the calculated structural ensembles.

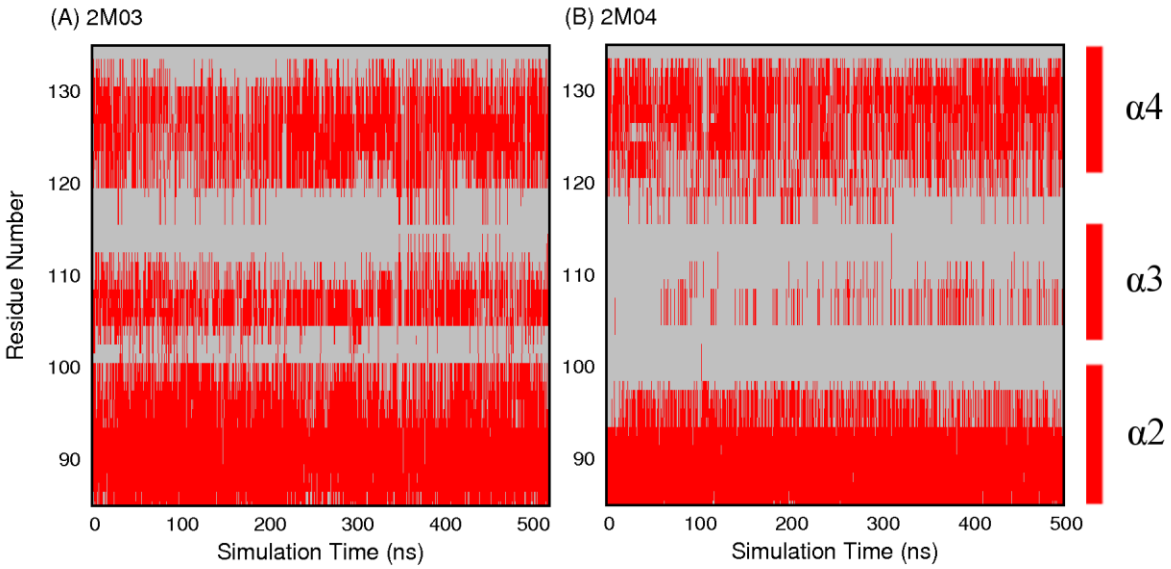


Figure 3.3. Helicity of residues near the BH3-only protein binding interface as a function of REST simulation time at condition λ = 1. Helical residues are colored in red.

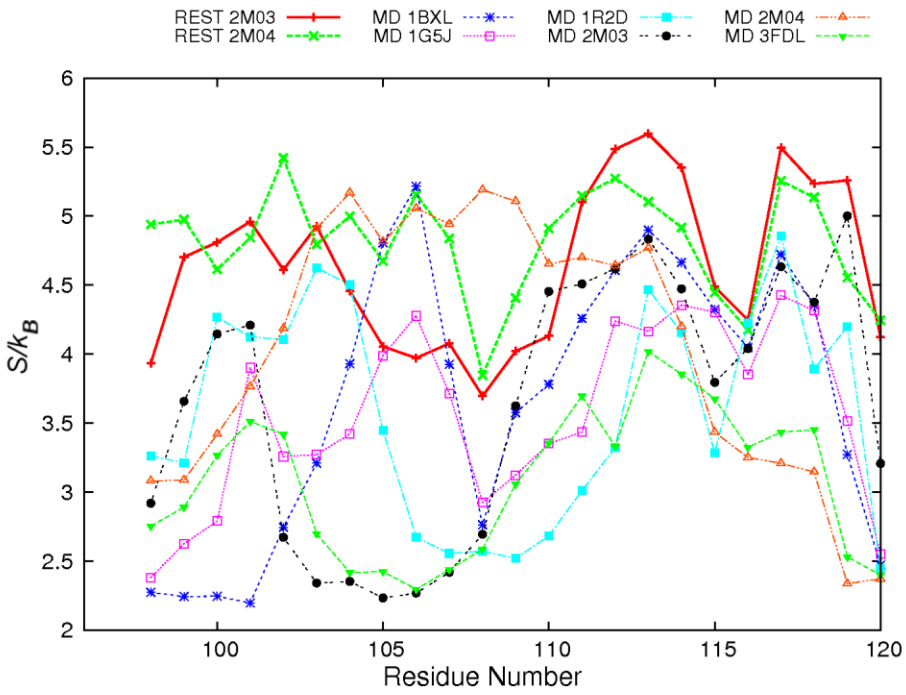


Figure 3.4. Residue backbone conformational entropy in the BH3-only protein binding interface of Bcl-xL. The first 300 ns data of REST simulations were discarded. Results from MD simulations are obtained by analyzing trajectories from our previous work[94].

3.3.2 Validation of REST-derived ensembles of unbound Bcl-xL

The reliability of the ensembles generated by REST simulations has been mainly examined by comparing the back-calculated NOE distances with experimental values. As shown in Table 3.2, all NOE distances for pairs of protons within the BH3-only protein binding interface region obtained from both REST simulations are within ~1 Å of experimentally assigned upper bounds. This is notable since the initial structures of these simulations are significantly different from each other (Figure 3.1). This implies the local structures of the Bcl-xL interface derived from these two REST simulations are likely realistic. Table 3.3 and Figure 3.12 summarize NOE violation analysis for pairs of protons between the BH3-only protein binding interface and the rest of the Bcl-xL protein. For the ensemble derived

from REST simulation 2M03, only 3 out of the 30 NOEs are violated by over 1 Å. On the other hand, the ensemble derived from simulation 2M04 yields more violations, with 14 out of 30 NOEs violated by at least 2 Å. This likely reflects the formidable challenge of sampling tertiary packing among different protein segments. Nonetheless, further analysis reveals that the contact maps between residues of the BH3-only protein binding interface region and the rest of the protein are actually similar from both ensembles (Figure 3.5). The implication is both REST simulations have sampled similar conformational space of the interface; however, the relative probabilities of various conformational substates are not quantitatively converged between two simulations. This notion is also consistent with PCA results showing a high level of overlap in conformational spaces sampled (Figure 3.2A and B). Importantly, both contact maps are highly consistent with the experimental NOE contact maps (black and red × in Figure 3.5). Taken together, although REST simulations were not fully converged and some NOE violations persist, the obtained ensembles appear to yield consistent local and tertiary structural features that are likely realistic.

Table 3.2. Summary of NOE violation analysis for proton pairs within the BH3-only protein binding interface region of Bcl-xL in the two REST simulations.

Proton Pairs		R_{NMR} (Å)	R_{2M03}^{REST} (Å)	Violation (Å)	R_{2M04}^{REST} (Å)	Violation (Å)
E98 H	L99 H	2.67	2.76	0.09	2.40	-
L99 H	R100 H	2.75	2.41	-	2.90	0.15
L99 H	Y101 H	3.69	4.09	0.40	4.71	1.02
R100 H	Y101 H	2.75	2.35	-	2.99	0.24
Y101 H	R102 H	2.96	3.10	0.14	2.81	-
R102 H	R103 H	4.53	2.57	-	2.65	-
R103 H	A104 H	2.85	3.14	0.29	3.33	0.48
A104 H	F105 H	3.10	3.24	0.14	3.24	0.14
A104 H	S106 H	4.36	4.77	0.41	5.40	1.04
F105 H	S106 H	3.27	2.40	-	3.00	-
F105 H	L108 H	4.74	4.90	0.16	5.66	0.92
S106 H	D107 H	3.79	2.60	-	2.25	-
D107 H	L108 H	3.10	2.30	-	2.70	-
L108 H	T109 H	2.92	2.62	-	2.72	-
L108 H	S110 H	3.58	4.10	0.52	4.25	0.67
T109 H	S110 H	3.13	2.59	-	2.44	-
T109 H	Q111 H	4.00	4.20	0.20	3.99	-
S110 H	L112 H	4.29	4.34	0.05	4.05	-
Q111 H	L112 H	2.85	2.48	-	2.30	-
Q111 H	H113 H	4.32	4.89	0.57	4.41	0.09
L112 H	H113 H	2.96	3.06	0.10	3.27	0.31
H113 H	I114 H	4.00	2.74	-	3.12	-
L99 H	L99 QQD	5.50	3.30	-	3.06	-
L99 QQD	R100 H	5.50	3.24	-	3.04	-
L99 QQD	Y101 H	5.50	4.83	-	4.47	-
L108 H	L108 QQD	4.73	3.25	-	2.97	-
L112 H	L112 QQD	5.50	3.12	-	3.13	-
L112 QQD	H113 H	5.50	2.75	-	2.91	-
I114 H	I114 QD1	5.17	3.11	-	3.32	-

R_{NMR} is experimentally determined NOE distance. R_{2M03}^{REST} and R_{2M04}^{REST} are calculated NOE-like distance derived from the two REST simulations.

Table 3.3. Summary of NOE violation analysis for proton pairs between the BH3-only protein binding interface (left column) and the rest of protein (right column) in the two REST simulations.

Atom Pairs		R_{NMR} (Å)	$R_{\mathrm{2M03}}^{\mathrm{REST}}$ (Å)	Violation (Å)	$R_{\mathrm{2M04}}^{\mathrm{REST}}$ (Å)	Violation (Å)
E98 H	E96 H	3.79	4.28	0.49	4.36	0.57
E98 H	F97 H	3.06	2.67	-	2.47	-
E98 H	V141 QQG	5.41	4.76	-	5.59	0.18
L99 H	F97 H	3.94	4.14	0.20	4.00	0.06
L108 QQD	V126 H	5.29	5.46	0.17	7.80	2.51
L108 QQD	V127 H	5.29	5.61	0.32	9.20	3.91
L108 QQD	E129 H	5.29	4.79	-	8.76	3.47
L108 QD1	L130 QD1	3.56	2.97	-	6.21	2.65
L108 QD1	L130 QD2	3.56	3.38	-	6.43	2.87
L108 QD2	L130 QD1	3.56	2.52	-	6.25	2.69
L108 QD2	L130 QD2	3.56	2.89	-	7.14	3.58
L108 QQD	L130 H	5.29	3.68	-	7.81	2.52
L108 QQD	L130 QQD	2.74	2.25	-	5.11	2.37
L108 QQD	F131 H	5.29	5.71	0.42	10.57	5.28
L108 QD1	F146 H	5.50	5.82	0.32	6.11	0.61
L108 QD2	F146 H	5.50	4.76	-	5.15	-
L108 QQD	G147 H	5.29	6.11	0.82	6.17	0.88
L108 QQD	L150 H	5.78	5.52	-	4.21	-
I114 H	L150 QQD	5.09	3.86	-	3.65	-
I114 QD1	L150 QQD	2.71	2.57	-	2.65	-
I114 QD1	S154 H	5.94	4.07	-	7.18	1.24
I114 QD1	V155 H	5.50	5.95	0.45	7.98	2.48
I114 QD1	L162 QD1	3.62	3.27	-	3.32	-
I114 QD1	L162 QD2	3.62	2.69	-	3.23	-
I114 QD1	V163 H	5.50	7.06	1.56	8.88	3.38
I114 QD1	S164 H	6.99	8.95	1.96	10.14	3.15
I114 QD1	R165 H	6.00	8.19	2.19	8.88	2.88
I114 QD1	I166 QD1	2.88	3.11	0.23	4.40	1.52
Y120 H	Q121 H	3.31	2.90	-	2.67	-
Y120 H	W169 HE1	3.79	3.74	-	4.89	1.10

R_{NMR} is experimentally determined NOE distance restraint. $R_{\mathrm{2M03}}^{\mathrm{REST}}$and $R_{\mathrm{2M04}}^{\mathrm{REST}}$ are calculated NOE-like distances for the two REST simulations.

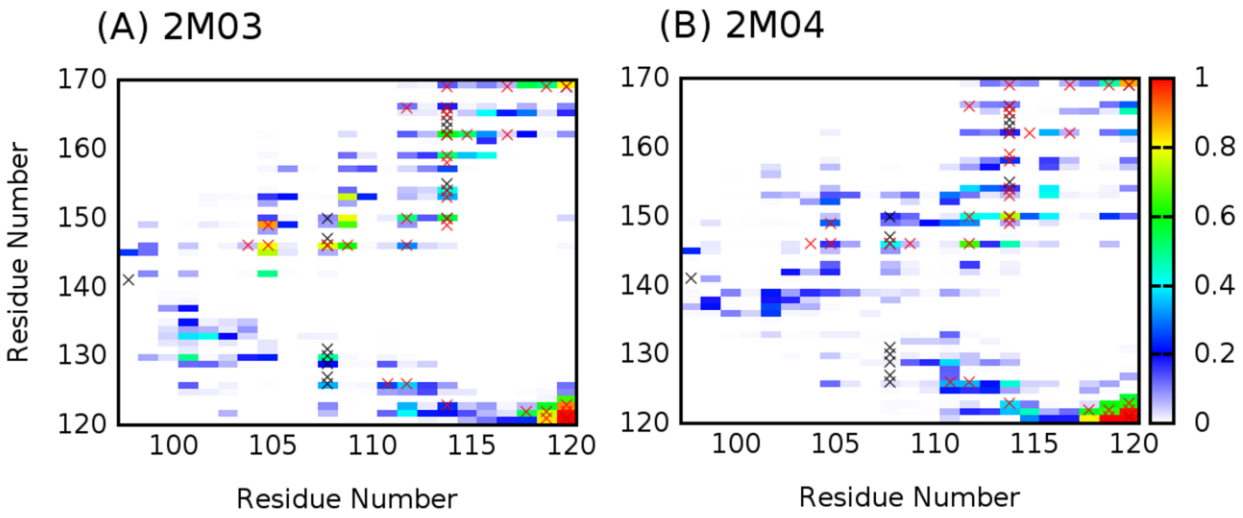


Figure 3.5. Contact probability maps between residues of the BH3-only protein binding interface region (residues 98-120) and the rest of the protein computed from the last 200 ns of REST simulations 2M03 and 2M04, respectively. Contact formation is defined when the minimal heavy-atom-distance between two residues is ≤ 4.2 Å. The probability of forming residue contacts in our simulations is shown in heat map, and contacts detected experimentally from NOE measurements are marked with black (from PDB 2M03) or red × (from PDB 2LPC).

3.3.3 Intrinsic flexibility and structural propensities of the BH3-only protein binding interface

The above analysis has strongly supported that enhanced sampling efficiency of REST has led to structural ensembles that are much better converged and more realistic. This allows us to further examine the conformational properties of the BH3-only protein binding interface of Bcl-xL to understand how they may provide a molecular basis for Bcl-xL recognition. Clearly, the interface is inherently flexible, with root-mean-square fluctuation (RMSF) values that are much higher than other helical regions and similar to terminal loops and tails (see Figure 3.6). We note that RMSF values of the interface region derived from REST run 2M04 are larger than those from run 2M03, which may be due to the unfolded initial structure (Figure 3.1) and thus over-representation of disordered conformations in the final ensemble.

Most interfacial residues are highly flexible and could sample both α-helix and extended conformations in the Ramachandra space (Figure 3.11A and B) and yield large backbone conformational entropy (Figure 3.4). This is also reflected in Figure 3.13 that solvent accessible surface area (SASA) of each interfacial residue shows large variance in both simulated ensembles and PDB ensembles. Moreover, packing between the BH3-only protein binding interface and the rest of the protein is not very tight, and many transient contacts could be formed between them (Figure 3.5). Interestingly, MD and NMR ensembles of unbound Bcl-xL have similar SASA profiles, despite substantial helical contents observed in MD. That is, residual structures do not lead to premature exposure of hydrophobic residues prior to binding (e.g., F105).

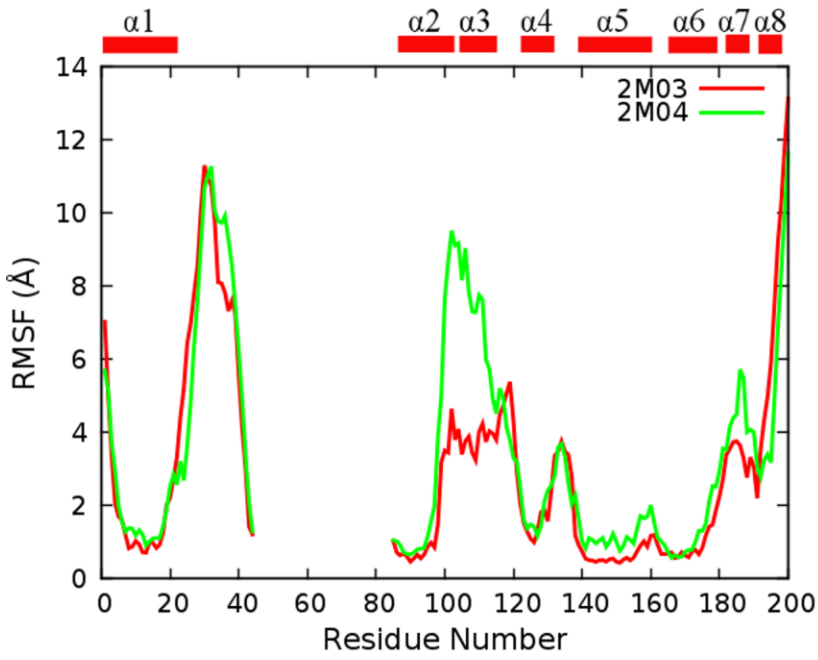


Figure 3.6. RMSF profiles of Bcl-xL Cα atoms calculated from REST trajectories (not including the first 300 ns data). All structures were superimposed based on Cα atoms of the core region (residues 85–98, 123–127, 140–156, and 162–175) before RMSF calculations.

The top eight most populated conformational "substates" identified from the clustering analysis are shown in Figure 3.7 and Figure 3.14, which further illustrate the highly heterogeneous nature of the BH3-only protein binding interface of Bcl-xL. The interfacial region could adopt various helical or disordered conformations, and interact with the rest of the protein through different packing modes. Moreover, based on our PCA analysis (Figure 3.2), which mainly captures the global motion of the interface regions, it's evident that all previously observed configurations could be visited by the free form of Bcl-xL. Although PC1 and PC2 are not necessarily appropriate reaction coordinates, energy barriers among various free energy basins seem to be mild (~2-3 kcal/mol), suggesting that conformational fluctuations at the BH3-only protein binding interface are likely rapid. For instance, both bound configurations of Bcl-xL in complex with BAX[145] (Figure 3.8A) or BAD[195] (Figure 3.8B) are well represented in the disordered ensemble derived from REST simulations. Although the existence of bound-like conformations in the unbound ensemble does not directly suggest conformational selection-like mechanism[26, 196-198] in Bcl-xL binding, accessibility of various bound states even in absence of ligands could facilitate Bcl-xL to respond rapidly to specific binding events, including the observed partial unfolding induced by PUMA binding.

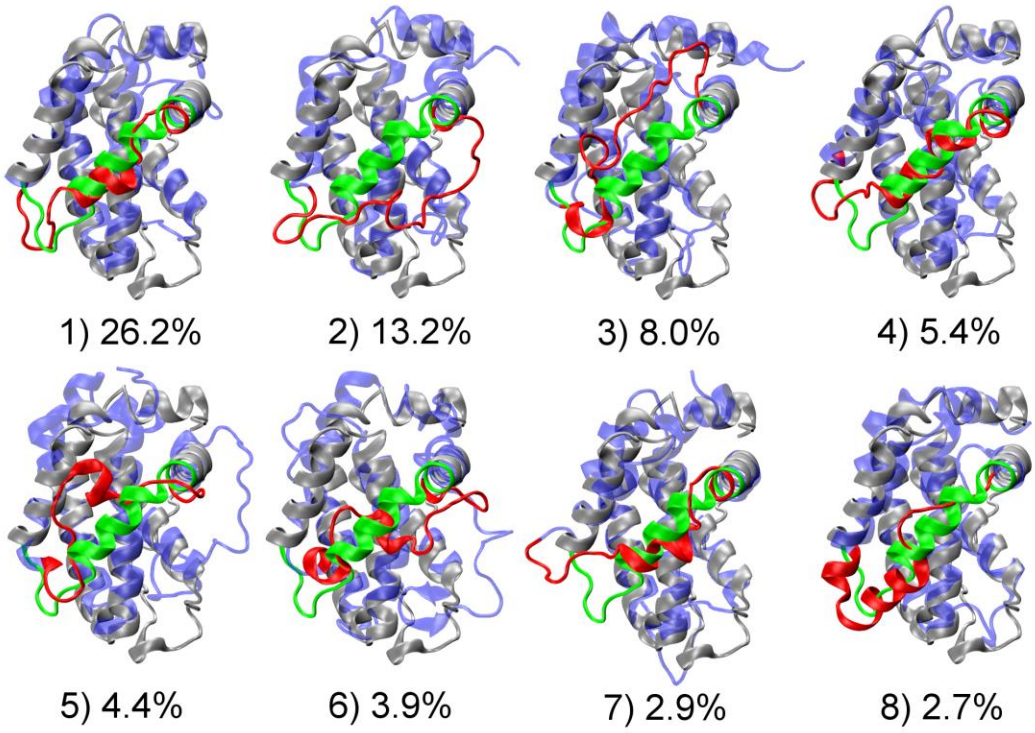


Figure 3.7. Central structures and populations of eight largest clusters for the simulated structure ensemble. Initial structure of simulation 2M03 is shown in grey cartoon, with the BH3-only protein binding interface highlighted in green. Central structure of each cluster is shown in blue, with the interfacial region colored in red. Clusters 1, 4, 7 and 8 are from 2m03 simulation, and 2, 3, 5 and 6 from 2m04 simulation.

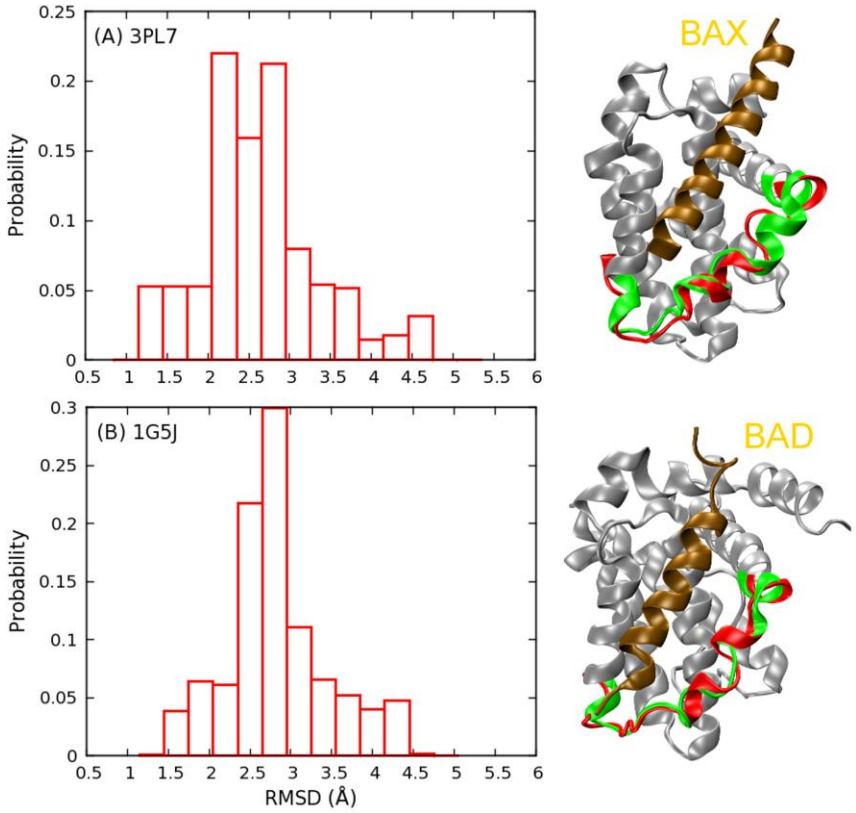


Figure 3.8. Probability distribution of interfacial Cα RMSD of the Bcl-xL structural ensemble derived from REST simulation 2M03 with respect to two PDB structures: (A) 3PL7 chain A (BAX bound form) and (B) 1G5J (BAD bound form). Representative low RMSD structures are shown. Bcl-xL in 3PL7 and 1G5J are colored in gray, with the BH3-only protein binding interface highlighted in green and bound BH3-peptide shown in ochre. Representative configurations of the interface sampled in REST are shown in red.

3.4 Conclusions

Previous NMR and biochemical studies have demonstrated that anti-apoptotic Bcl-xL undergoes local unfolding upon specific binding to the BH3-only protein PUMA, which in turn disrupts the interaction between Bcl-xL and tumor suppressor p53 and activates apoptosis[181]. However, the detailed mechanisms of such regulated unfolding of Bcl-xL have not been fully understood yet. Our previous studies have suggested that substantial intrinsic structure heterogeneity exists at the BH3-only protein binding interface of Bcl-xL[94], which could provide a general mechanism for supporting coupled binding, folding and unfolding in Bcl-xL recognition. We have leveraged enhanced sampling capability of REST to further characterize the conformational properties of unbound Bcl-xL at atomistic level in explicit solvent. Independent REST simulations initiated from two contrasting initial PDB structures yield structural ensembles that are not only much better converged but also highly consistent with available NMR data. The new simulations further support the notion that the binding interface of Bcl-xL has significant and inherent conformational heterogeneity. Intriguingly, the new results reveal that all previously observed conformations of Bcl-xL, including both bound and unbound states, are well represented in the unbound ensemble of Bcl-xL, and their interconversion appear to be rapid and involve small free energy barriers. Altogether, it is highly plausible that such inherent structural heterogeneity and plasticity at the binding interface is crucial for Bcl-xL to undergo rapid structural rearrangement and respond sensitively to various binding ligands for downstream signaling. Such a general molecular

mechanism may also be relevant for understanding other emerging cases of regulated unfolding in cellular signaling and regulation[107].

3.5 Supporting Material

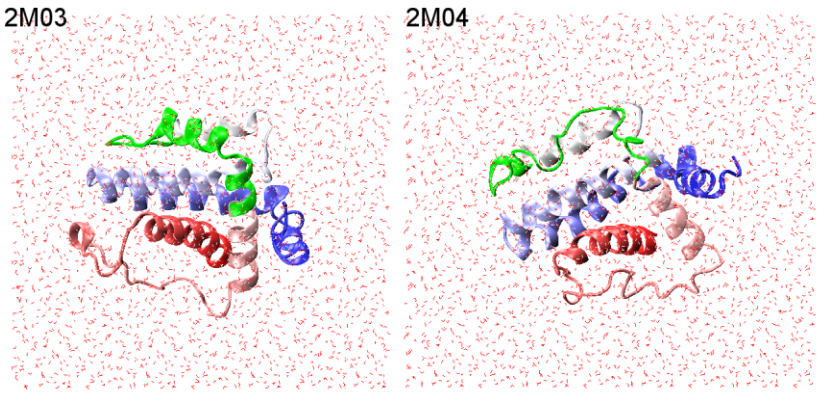


Figure 3.9 Initial structures of REST simulations. Water molecules are shown in red dots, and Bcl-xL in cartoon representation with color changing from red (at N-terminus) to blue (at C-terminus). The BH3-only protein binding interface (residues 98-120) is highlighted in green.

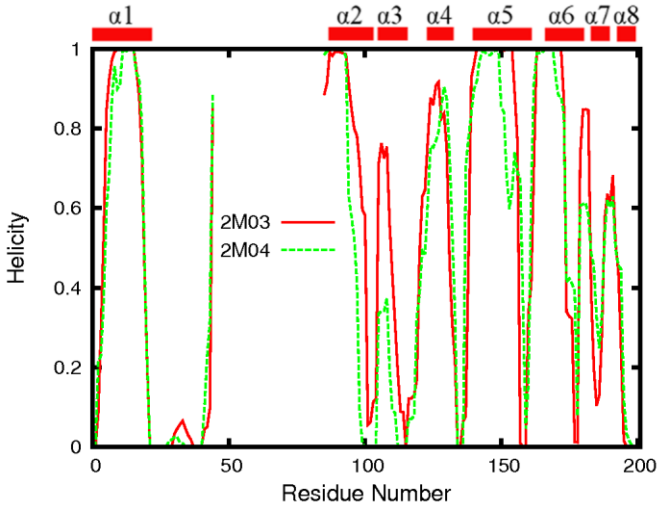
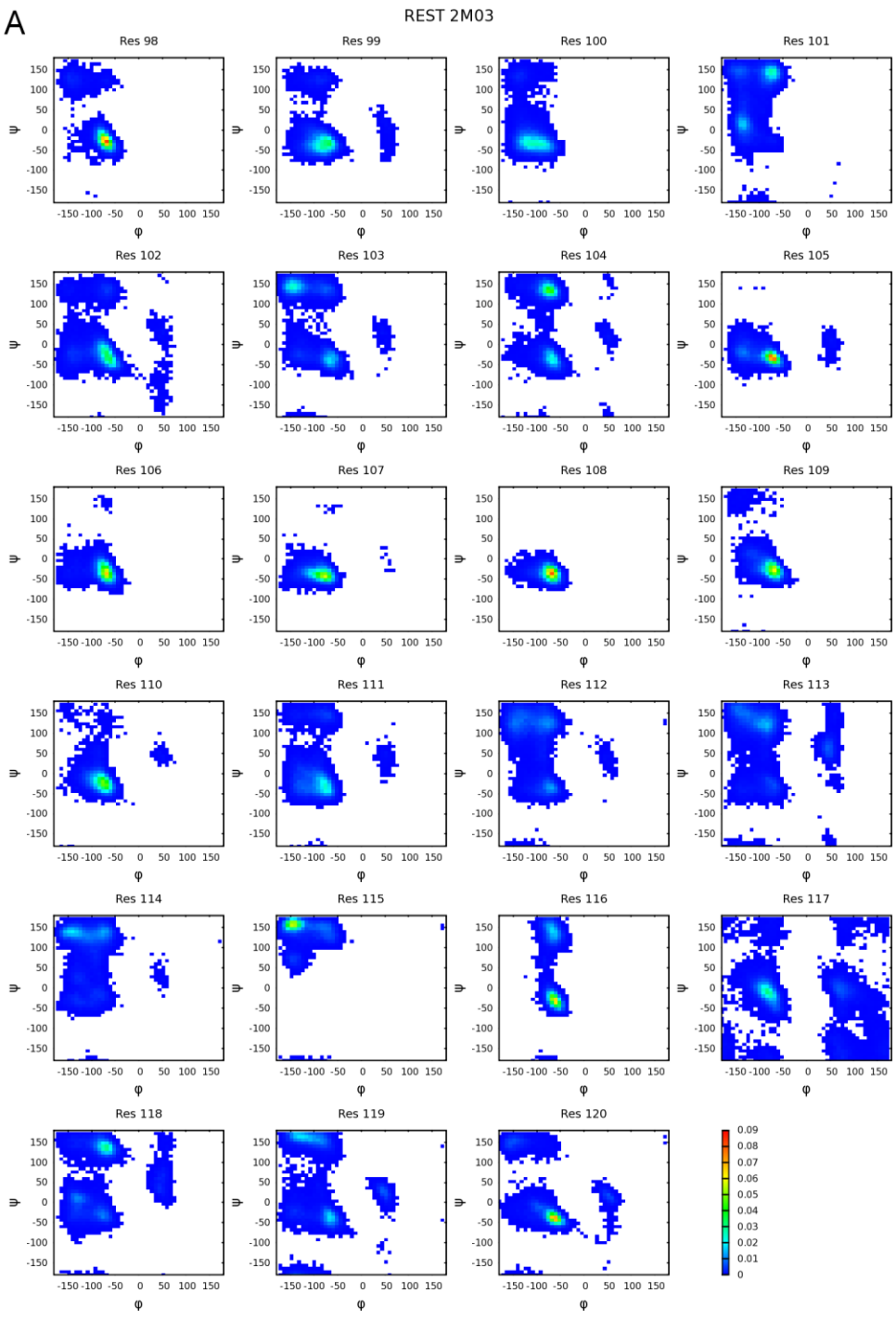
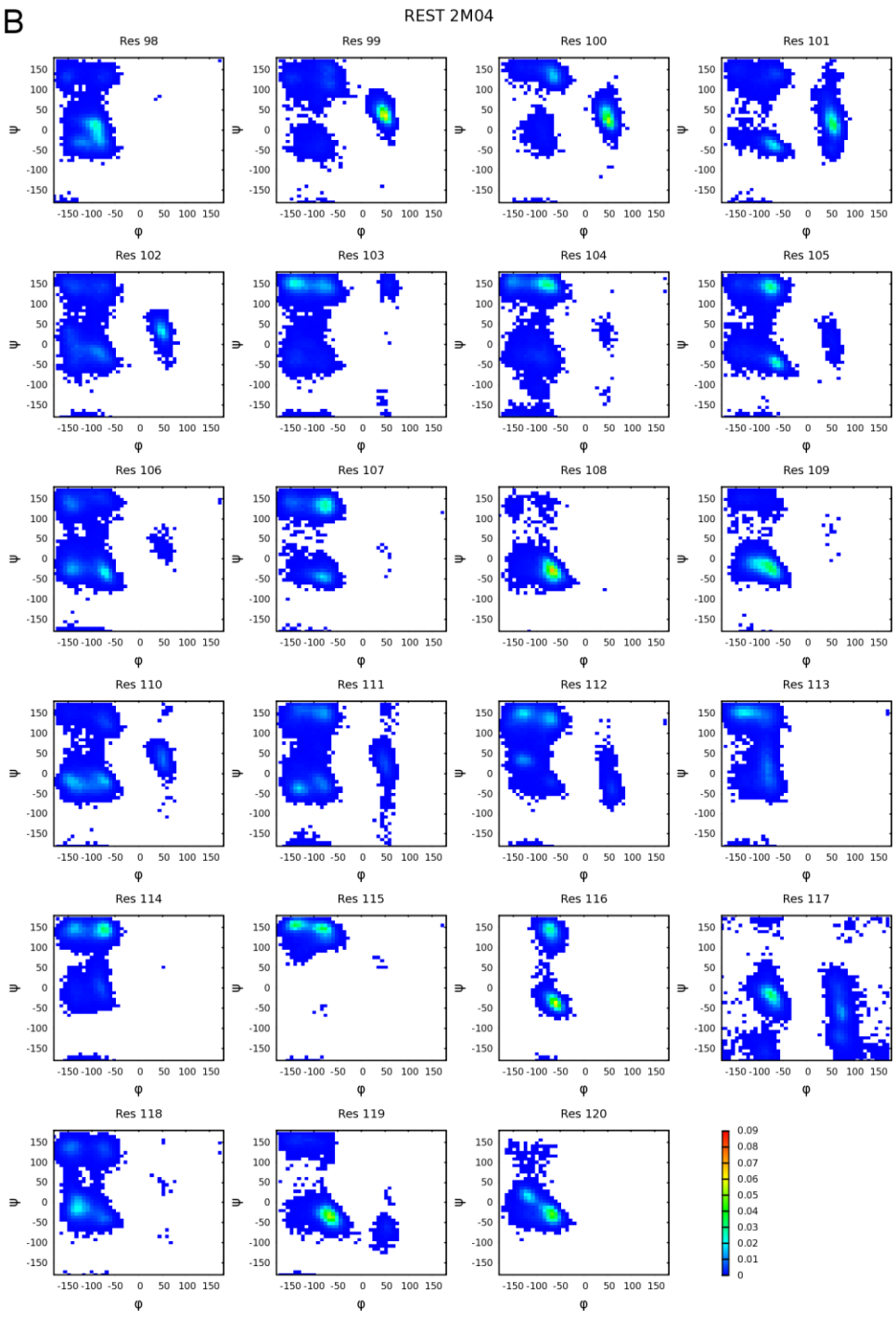
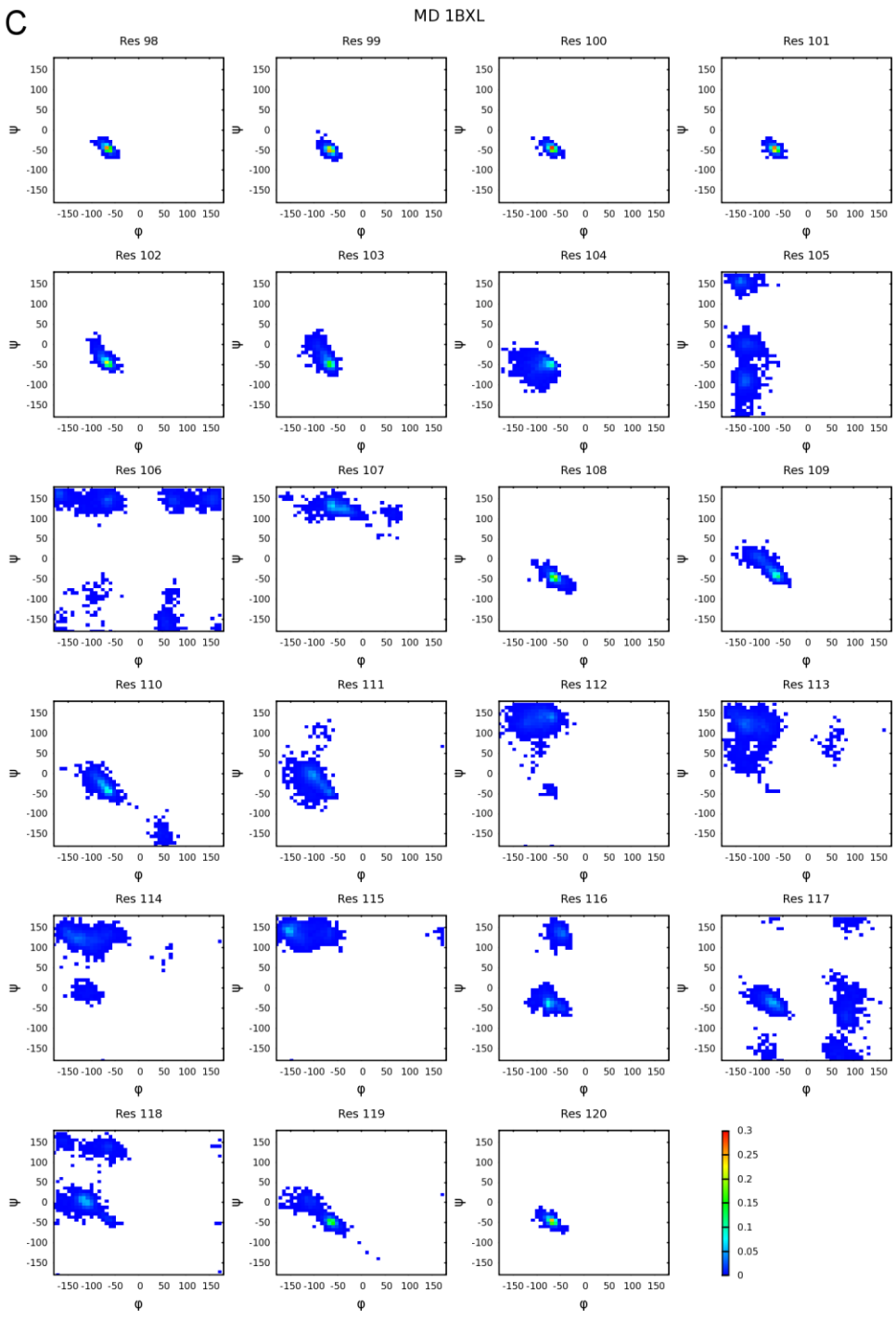
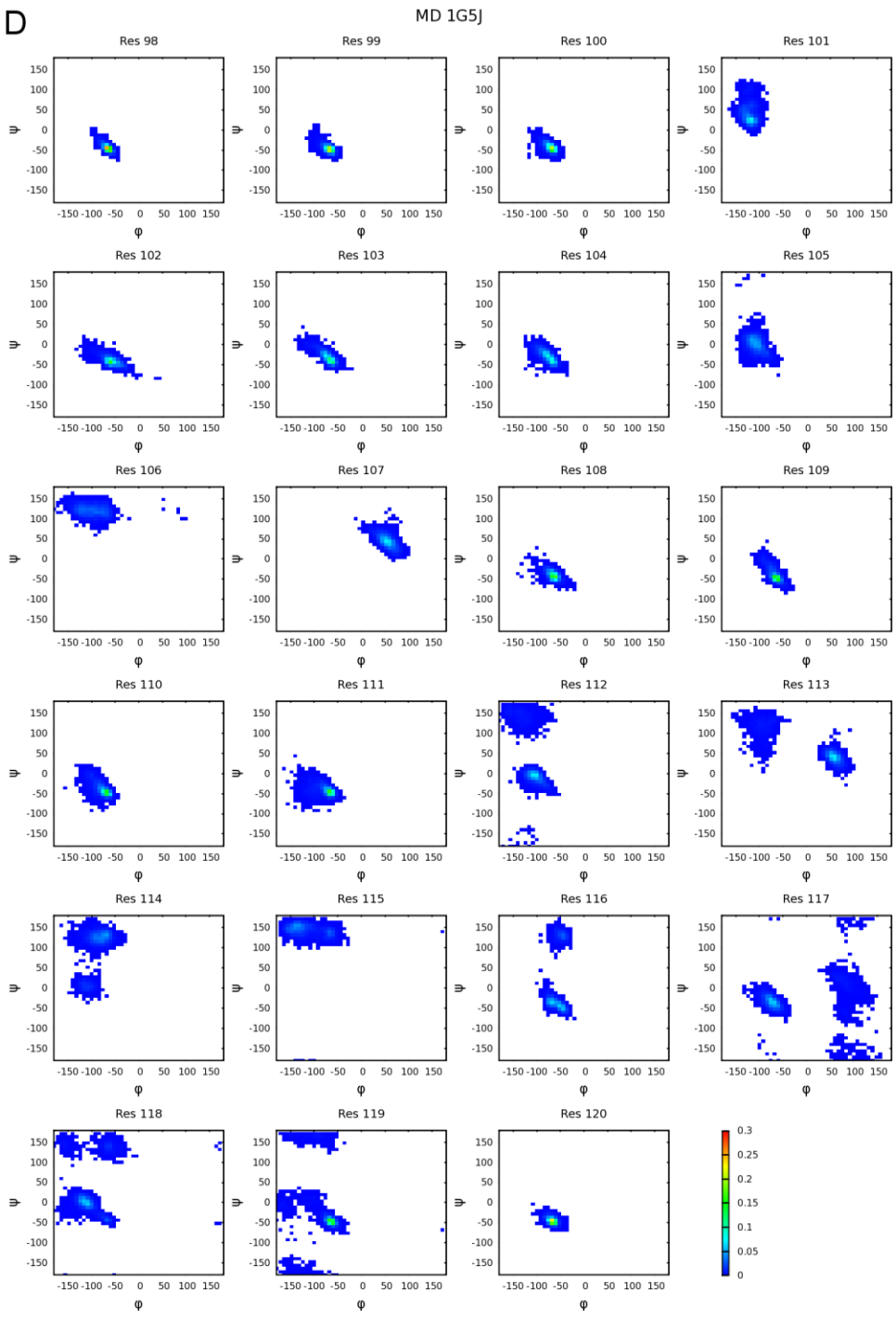


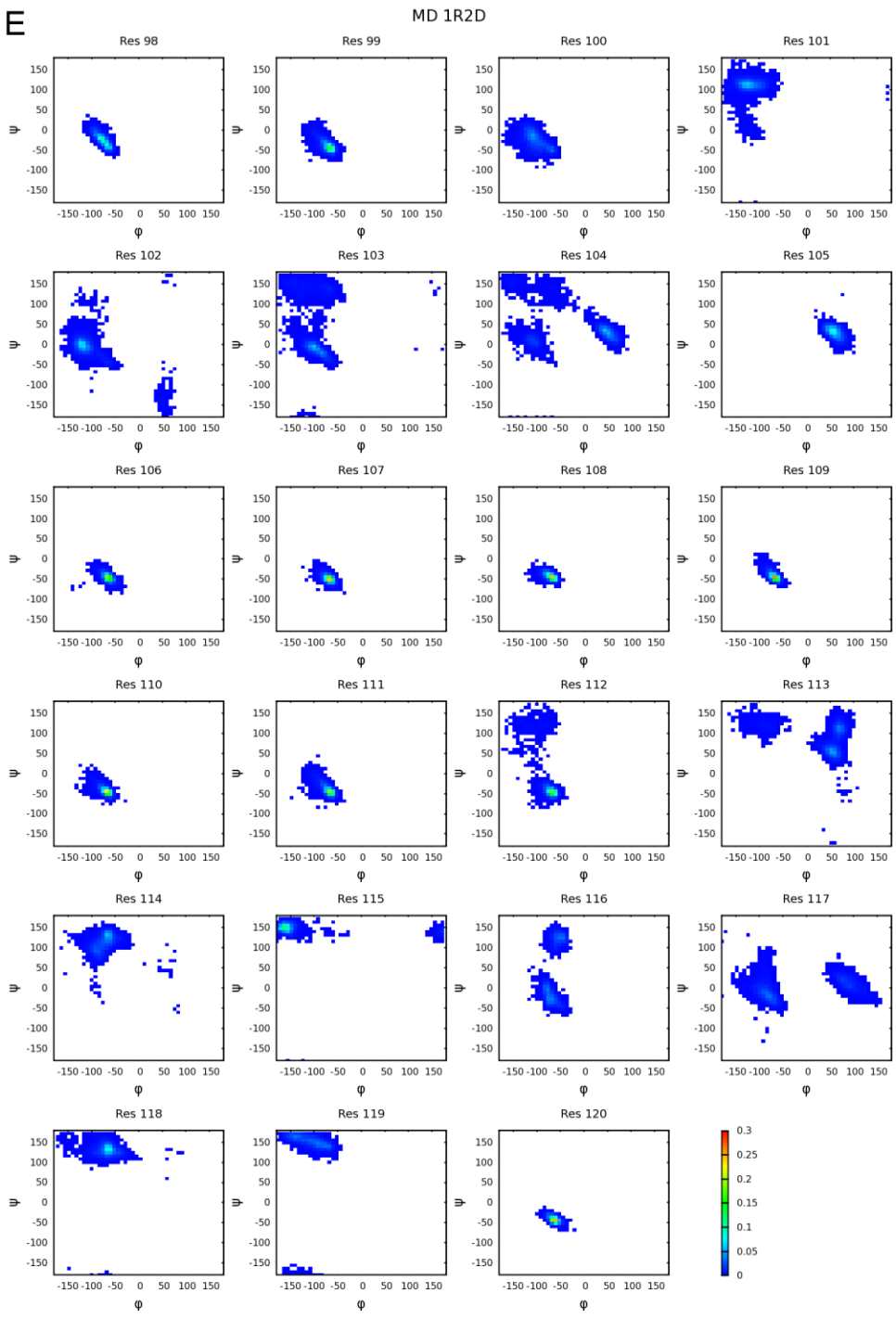
Figure 3.10. Averaged helicity profiles of Bcl-xL from two REST simulations. Only the last 200 ns data were used in this analysis.


A
REST 2M03
Res 98
Res 99
Res 100
Res 101
Res 102
Res 103
Res 104
Res 105
Res 106
Res 107
Res 108
Res 109
Res 110
Res 111
Res 112
Res 113
Res 114
Res 115
Res 116
Res 117
Res 118
Res 119
Res 120
ψ
φ
150
100
50
0
-50
-100
-150
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0

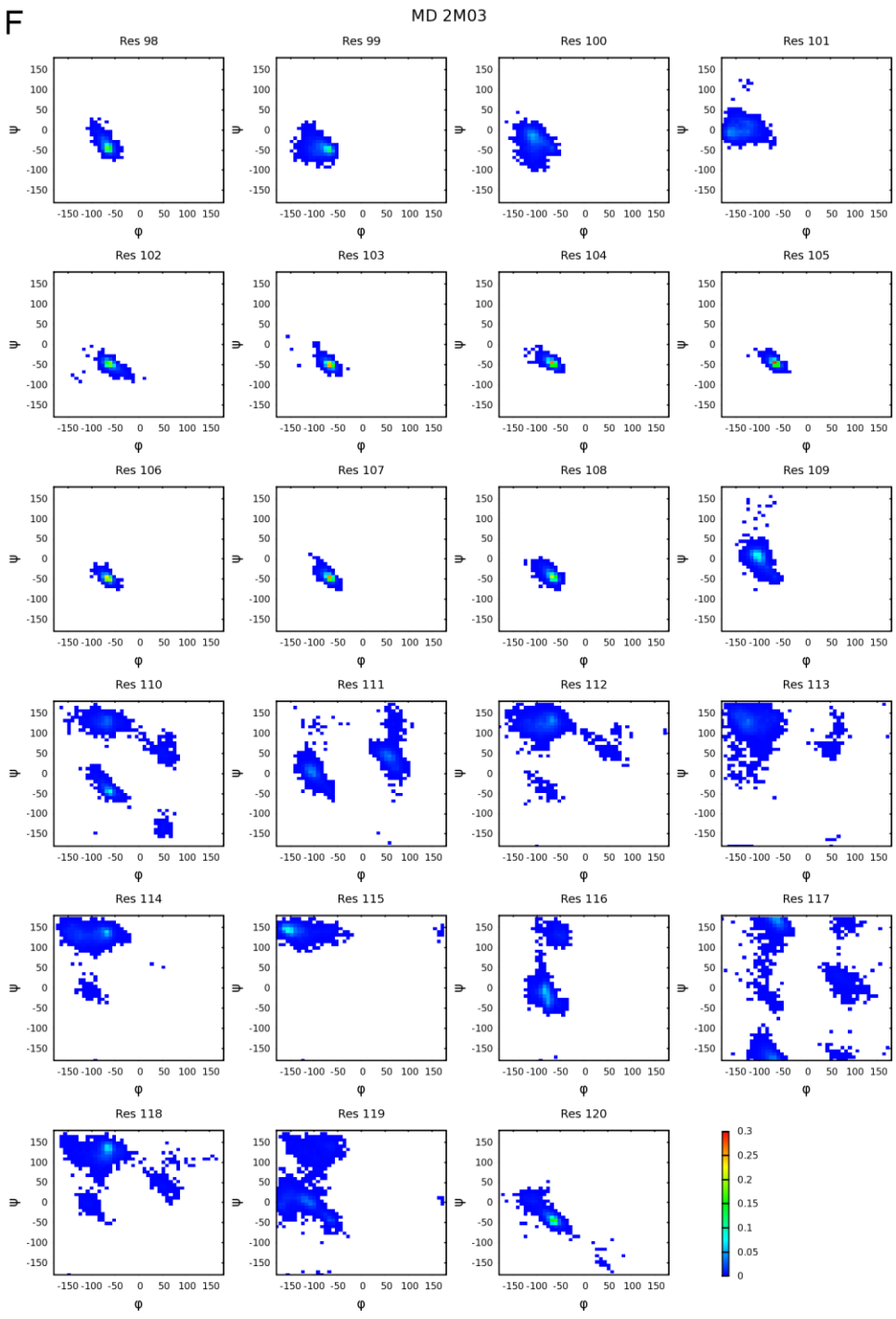

B
REST 2M04
Res 98
Res 99
Res 100
Res 101
Res 102
Res 103
Res 104
Res 105
Res 106
Res 107
Res 108
Res 109
Res 110
Res 111
Res 112
Res 113
Res 114
Res 115
Res 116
Res 117
Res 118
Res 119
Res 120
ψ
φ
150
100
50
0
-50
-100
-150
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0


C
MD 1BXL
Res 98
Res 99
Res 100
Res 101
Res 102
Res 103
Res 104
Res 105
Res 106
Res 107
Res 108
Res 109
Res 110
Res 111
Res 112
Res 113
Res 114
Res 115
Res 116
Res 117
Res 118
Res 119
Res 120
ψ
φ
150
100
50
0
-50
-100
-150
0.3
0.25
0.2
0.15
0.1
0.05
0

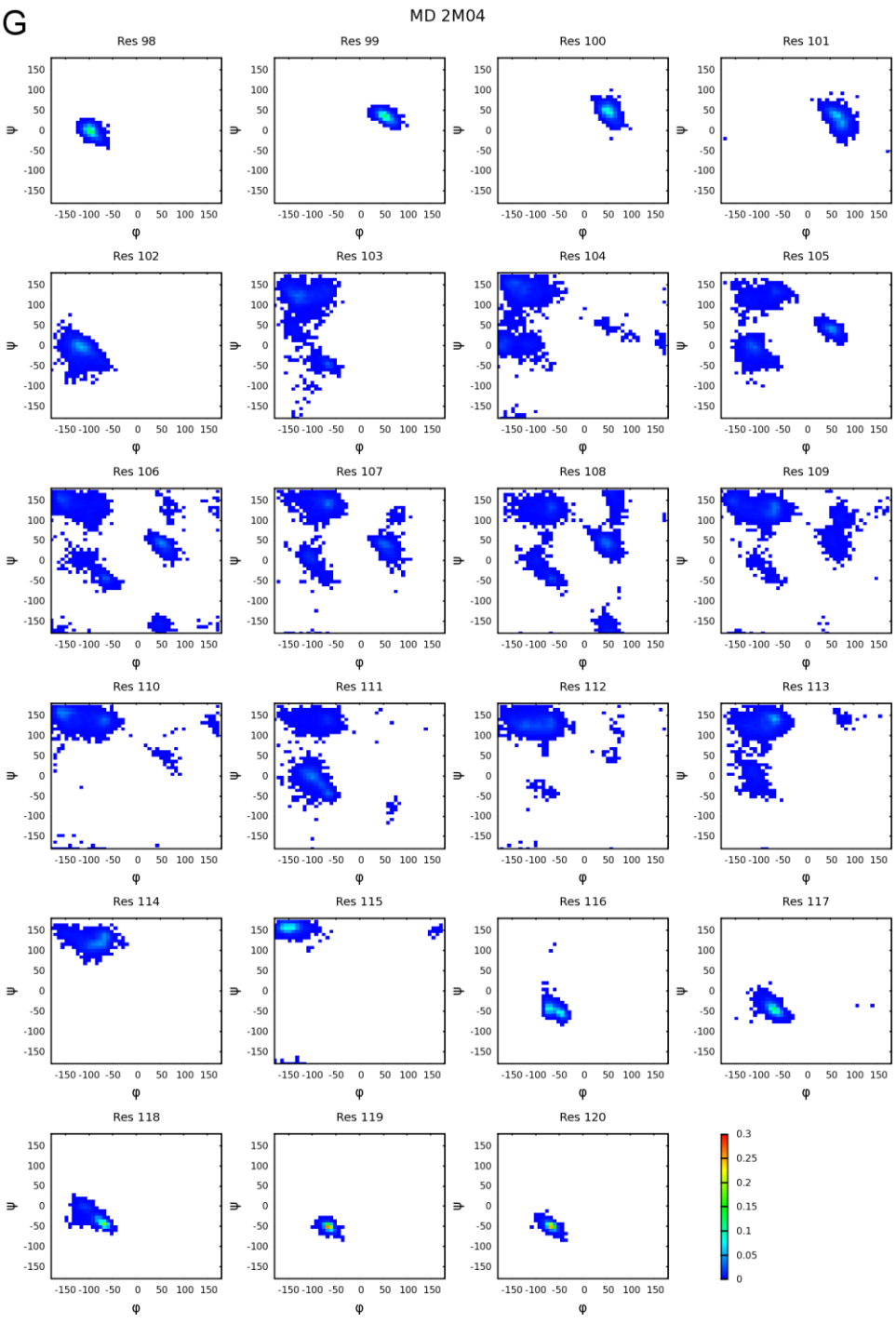

D
MD 1G5J
Res 98
Res 99
Res 100
Res 101
Res 102
Res 103
Res 104
Res 105
Res 106
Res 107
Res 108
Res 109
Res 110
Res 111
Res 112
Res 113
Res 114
Res 115
Res 116
Res 117
Res 118
Res 119
Res 120
ψ
φ
150
100
50
0
-50
-100
-150
0.3
0.25
0.2
0.15
0.1
0.05
0


E
MD 1R2D
Res 98
Res 99
Res 100
Res 101
Res 102
Res 103
Res 104
Res 105
Res 106
Res 107
Res 108
Res 109
Res 110
Res 111
Res 112
Res 113
Res 114
Res 115
Res 116
Res 117
Res 118
Res 119
Res 120
ψ
φ
150
100
50
0
-50
-100
-150
0.3
0.25
0.2
0.15
0.1
0.05
0

F


MD 2M03
Res 98
Res 99
Res 100
Res 101
Res 102
Res 103
Res 104
Res 105
Res 106
Res 107
Res 108
Res 109
Res 110
Res 111
Res 112
Res 113
Res 114
Res 115
Res 116
Res 117
Res 118
Res 119
Res 120
ψ
φ
150
100
50
0
-50
-100
-150
0.3
0.25
0.2
0.15
0.1
0.05
0

G


MD 2M04
Res 98
Res 99
Res 100
Res 101
Res 102
Res 103
Res 104
Res 105
Res 106
Res 107
Res 108
Res 109
Res 110
Res 111
Res 112
Res 113
Res 114
Res 115
Res 116
Res 117
Res 118
Res 119
Res 120
ψ
φ
150
100
50
0
-50
-100
-150
0.3
0.25
0.2
0.15
0.1
0.05
0

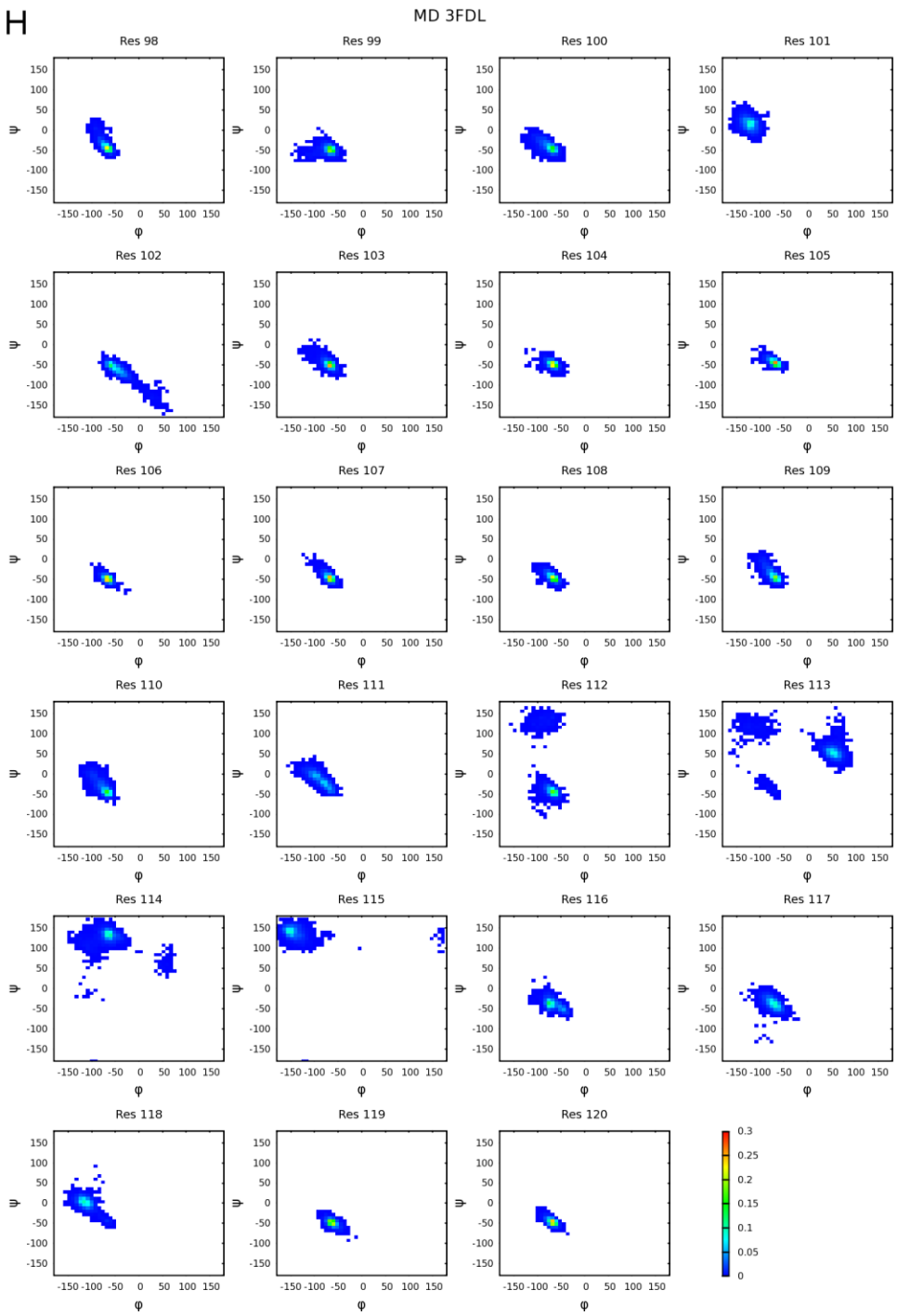


Figure 3.11 Two-dimensional probability distribution of (φ, ψ) for residues at the Bcl-xL interface region from two REST simulations (this work) and six MD simulations (from the previous work, see Ref 29 of the main text).

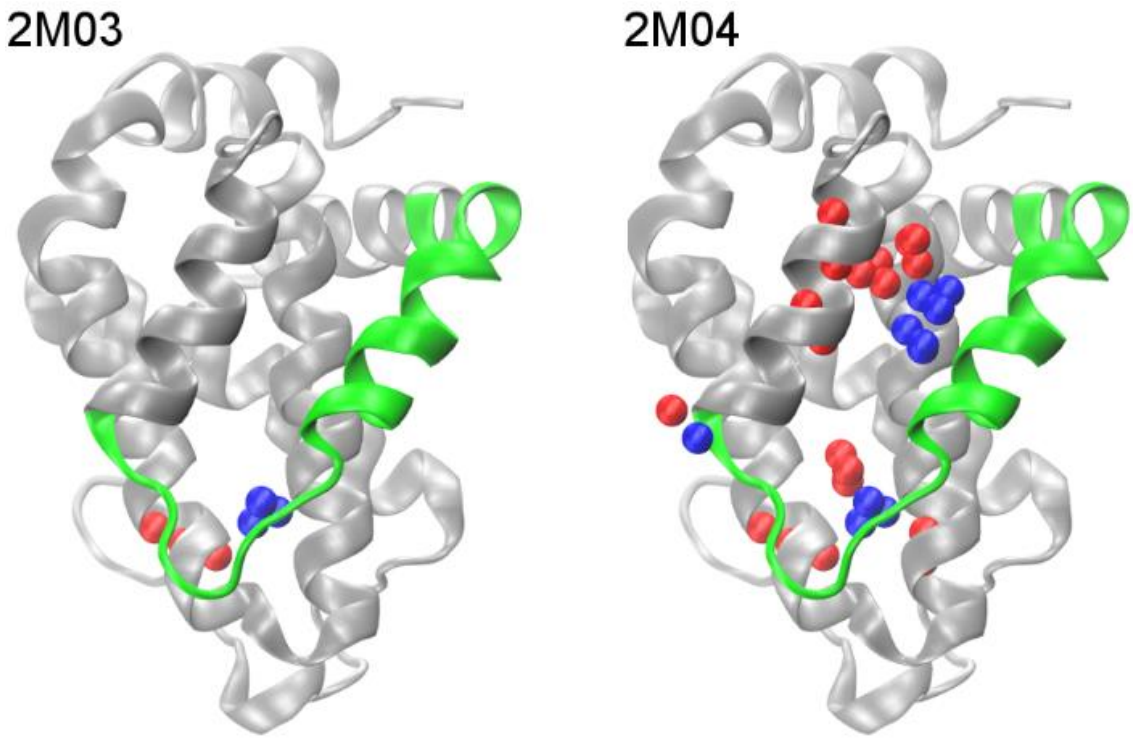


Figure 3.12 Mapping of NOE violations greater than 1 Å to PDB structure 2M03. All NOE distances for pairs of protons within the BH3-only protein binding interface region obtained from both REST simulations are within ~1 Å of experimentally assigned upper bounds. Therefore, only proton pairs between the BH3-only protein binding interface and the rest of protein were shown here for REST 2M03 (left) and 2M04 (right) simulations. Bcl-xL is in gray cartoon, with the BH3-only protein binding interface highlighted in green. Protons within the BH3-only protein binding interface and the rest of protein are colored in blue and red, respectively.

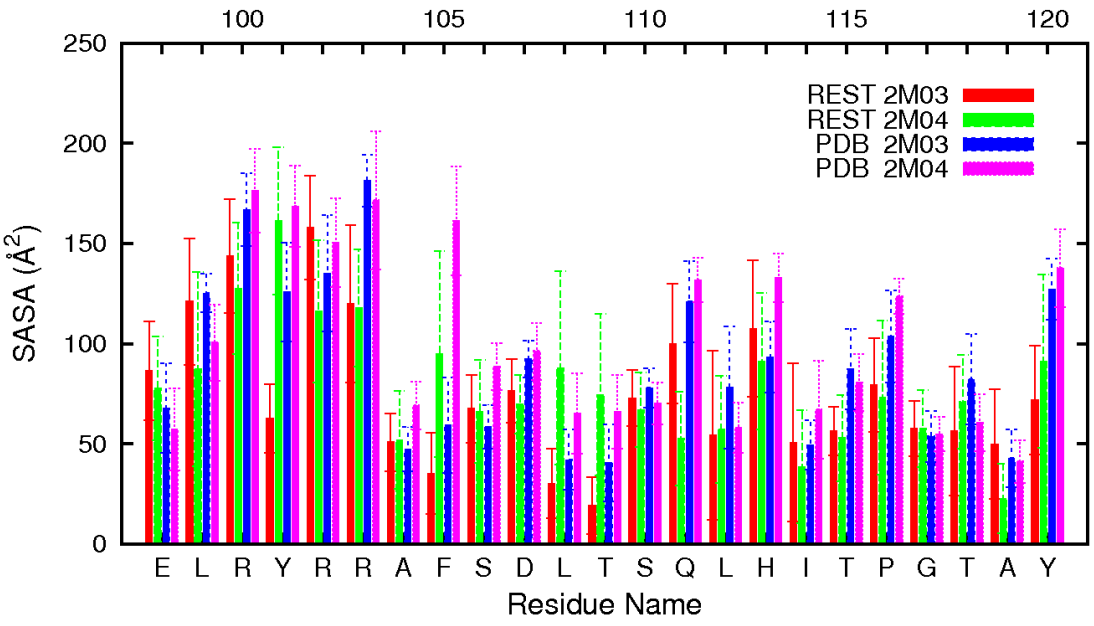


Figure 3.13 Mean and standard deviation of solvent accessible surface area (SASA) for each residue in the BH3-only protein binding interface of Bcl-xL calculated from simulated and PDB ensembles. For simulated ensembles, the first 300 ns trajectories were excluded in the calculation. For PDB ensembles, all models in PDB 2M03 (unbound state) or 2M04 (bound state) were analyzed. The bound PUMA in 2M04 was removed from before analysis. A water probe with 1.4 Å radius was used in all SASA calculations.

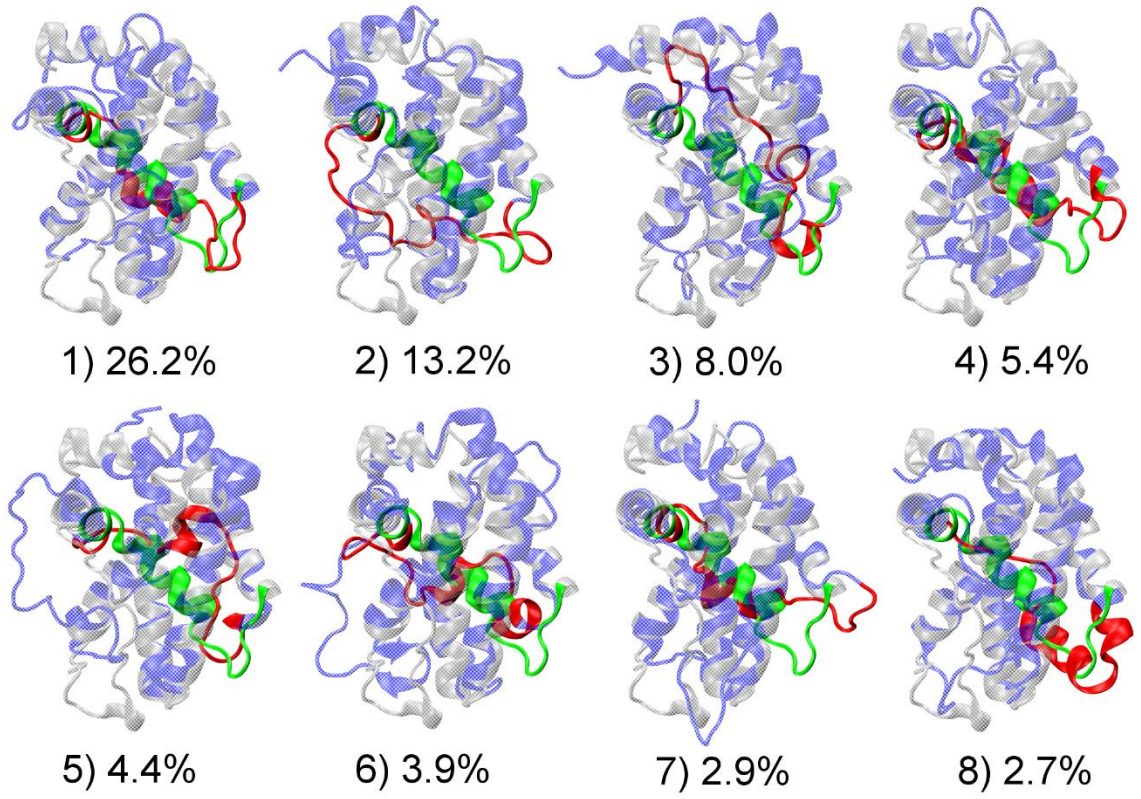


Figure 3.14 Back view of central structures and populations of eight largest clusters for the simulated structure ensemble. Initial structure of simulation 2M03 is shown in grey cartoon, with the BH3-only protein binding interface highlighted in green. Central structure of each cluster is shown in blue, with the interfacial region colored in red.

CHAPTER 4

RESIDUAL STRUCTURE ACCELERATES BINDING OF INTRINSICALLY DISORDERED ACTR BY PROMOTING EFFICIENT FOLDING UPON ENCOUNTER*

Intrinsically disordered proteins (IDPs) often fold into stable structures upon specific binding. The roles of residual structure of unbound IDPs in coupling binding and folding have been under much debate. While many studies emphasize the importance of conformational flexibility for IDP recognition, it was recently demonstrated that stabilization the N-terminal helix of intrinsically disordered ACTR accelerated its binding to another IDP, NCBD of the CREB-binding protein. To understand how enhancing ACTR helicity accelerates binding, we derived a series of topology-based coarse-grained models that mimicked various ACTR mutants with increasing helical contents and reproduced their NCBD binding affinities. Molecular dynamics simulations were then performed to sample hundreds of reversible coupled binding and folding transitions. The results show that increasing ACTR helicity does not alter the baseline mechanism of synergistic folding, which continues to follow "extended conformational selection" with multiple stages of selection and induced folding. Importantly, these coarse-grained models, while only calibrated based on binding thermodynamics, recapitulate the observed kinetic acceleration with increasing ACTR helicity. However, the residual helices do not enhance the association kinetics via more efficient seeding of

* Liu X, Chen J and Chen J (2018). "Residual Structures Accelerate Binding of Intrinsically Disordered ACTR by Promoting Efficient Folding upon Encounter." Journal of Molecular Biology 431(2): 422-432

productive collisions. Instead, they allow the nonspecific collision complexes to evolve more efficiently into the final bound and folded state, which is the primary source of accelerated association kinetics. Meanwhile, reduced dissociation kinetics with increasing ACTR helicity can be directly attributed to less entropic penalty of forming the bound state. Altogether, this study provides important mechanistic insights into how residual structure may modulate thermodynamics and kinetics of coupled binding and folding of IDPs.

4.1 Introduction

Unlike well-folded proteins, intrinsically disordered proteins (IDPs) lack stable 3D structures in the unbound state under physiological conditions[1-3, 21, 53, 98, 170, 199, 200]. They play important roles in cellular protein-protein interaction networks and are capable of interacting with many targets with specificity[2, 3, 13, 17, 100, 169, 170]. Upon specific binding, IDPs often gain stable secondary and/or tertiary structures[100, 161, 179]. The molecular mechanism of how IDPs achieve efficient coupled binding and folding has been under intensive studies[61, 81, 161, 198, 201-205]. In particular, the unbound states of IDPs often contain residual structure that resembles those in the bound state[25, 53, 81, 100, 199, 206, 207], even though the roles of such residual structure in IDP recognition remain debatable. On one hand, residual structure may serve as initial contact points that facilitate productive binding and folding, referred to as conformational selection-like mechanisms[25]. On the other hand, conformational flexibility has been argued to be crucial for binding, and increasing the level of residual structure may reduce the association

kinetics[24, 70, 208]. It has been argued that rapid folding upon encounter is critical for IDPs to achieve facile specific recognition[209, 210]. The dual-transition-state theory[211] predicts that the diffusion-limited encounter rate represents an upper-limit for that of IDP coupled binding and folding, which cannot be achieved unless IDPs fold rapidly upon encounter (i.e., beyond the typical speed limit of μs^{-1} for isolated proteins[212]). Curiously, it has been observed that IDPs have similar association rates compared to folded proteins[213, 214]. The implication is that IDPs could fold efficiently upon encounter in general. Several features of IDPs have been argued to contribute to efficient folding upon encounter, such as small interacting domains, simple folded topologies with low contact orders, and likely an appropriate balance between residual structure and conformational flexibility[53].

In a recent NMR and stopped-flow kinetic study, Kjaergaard and coworkers examined the effects of stabilizing residual helices on the association of activation domain of the activator for thyroid hormone and retinoid receptors (ACTR) with the nuclear coactivator binding domain (NCBD) of CREB-binding protein[31] (Figure 4.1). Eight ACTR mutants were designed with varying helical propensities in the N-terminal helix (H1), without perturbing its electrostatic properties or inter-molecular interactions. Intriguingly, increased helicity of ACTR H1 was observed to accelerate the rate of association with NCBD and at the same time decelerate the dissociation rate, leading to a net stabilization of the complex. Such accelerated association induced by increasing residual helicity has also been reported for the association of KIX with c-Myb[215, 216] and assembly of the spectrin tetramerization domain[217]. Yet, this observation is in contrast to several previous studies where

the association rates are either reduced (e.g., in $p27^{Kip1}$/cyclin A/Cdk2[208] and pKID/KIX[70] interactions) or remain similar (e.g., PUMA binding to MCL-1[203]) with increasing residual helicity. Mechanistically how stabilized residual helices modulate the interaction kinetics and mechanism is not clear. As discussed in the previous work [31], one plausible explanation is that the rate-limiting folding step occurs after an initial binding step, and the energy barrier decreases with increased helical content[31]. Another interpretation could be the existence of several parallel pathways, and increasing helical population may significantly increase the flux of conformational selection like pathways (rather than lower the barrier height)[31].

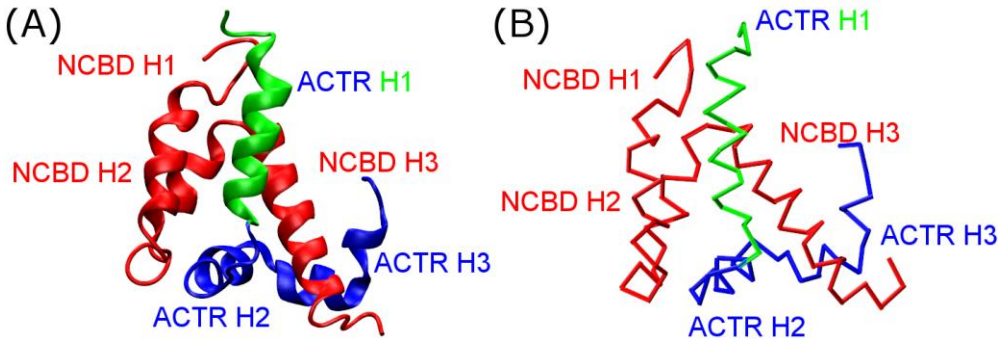


Figure 4.1(A) Structure of the NCBD/ACTR complex (PDB 1KBH[218]) and (B) the Cα-only Gō-like model.

Molecular dynamics (MD) simulations could provide microscopic details necessary for unveiling the molecular mechanisms of complex coupled binding and folding processes, which often involve multiple intermediate states and parallel pathways[61, 70, 81, 197, 198, 210, 219]. However, atomistic simulations using physics-based force fields remain computationally too expensive to sample reversible binding and folding transitions to obtain reliable and statistically meaningful observations on mechanism[53]. Instead, topology-based coarse-grained (CG) modeling has been shown to offer an effective tool for mechanistic

studies of coupled binding and folding of IDPs into stable complexes [70, 219, 220], which arguably is also governed by the minimal frustration principle of protein folding[221]. With proper calibration to balance the interplay of residual folding and intermolecular interactions, these Gō-like models have been successfully applied to several IDP complexes, and many predictions have been substantiated by independent experiments [70, 198, 209, 222-224]. In this work, we first derive a series of Gō-like CG models that are carefully calibrated to mimic ACTR mutants with various residual helical contents and reproduce their NCBD binding affinities. Milliseconds of MD simulations were then performed to sample hundreds of reversible coupled binding and folding transitions to analyze the association kinetics and mechanism. The results show that these Gō-like models, while only calibrated based on binding thermodynamics, recapitulate the observed kinetic acceleration with increasing ACTR helicity. Mechanistic analyses reveal that pre-existing structures in ACTR do not significantly alter the baseline mechanism of its synergistic folding upon binding to NCBD. Instead, they accelerate the overall association kinetics mainly by promoting more efficient folding upon encounter.

4.2 Results and Discussion

4.2.1 Gō-like models recapitulate higher NCBD/ACTR affinity with stabilized ACTR H1

As summarized in Figure 4.2A and Table 4.1, scaling of the intra-molecular interaction strengths of ACTR H1 allows direct modulation of its average helicity. The average helicity is bound between 10% to 95% due to the coarse-grained nature of Gō-like model. However, the resulting models overestimate how much

NCBD/ACTR is stabilized by increasing ACTR H1 helicity, probably because the coarse-grained models often could not properly balance entropy and enthalpy, and directly stabilizing ACTR H1 overly reduce the entropic cost of binding. To capture the binding affinities of ACTR to NCBD measured experimentally[31], the inter-molecular interaction strength between ACTR and NCBD needs to be slightly scaled down, which was determined by replica exchange (REX) simulations in combination with Hamiltonian mapping[225, 226]. As shown in Figure 4.2B, $\log(K_d)$ of the final calibrated models shows a strong correlation (R^2 = 0.98) with $\log(K_{helix})$, as observed experimentally[31]. The slopes are also similar for the simulated and experimental values[31] (-0.80 and -0.84, respectively). Consistently, the free energy profiles as a function of the fraction of total inter-molecular native contacts (Q_{inter}), presented in Figure 4.3A and B, illustrate that the bound state becomes increasingly favorable as ACTR gains more helical structures at H1 region. Experimentally, various ACTR mutants were designed without perturbing its electrostatic interactions or inter-molecular interactions in the complex[31]. We can assume that the bound state remains similar in all mutants. Therefore, increasing ACTR secondary structures could reduce the overall entropic cost of forming NCBD/ACTR complex, to result in augmented complex stability. By decomposing the free energy into enthalpic and entropic contributions (Figure 4.9), we found that the overall entropic penalty ($T\Delta S$) from unbound state (Q_{inter} = 0) to fully bound state (Q_{inter} = 0.6) is indeed reduced from 78.3 kcal/mol to 73.4 kcal/mol as we increase ACTR H1 helicity from 0.44 to 0.80 (Figure 4.9B upper inset). Being able to capture the key features of experimentally observed structural and thermodynamic

properties, the calibrated Gō-like models should enable us to investigate how residual helicity of ACTR H1 modulates the binding kinetics and pathways of NCBD/ACTR interaction.

Table 4.1 Summary of production simulation details and derived NCBD/ACTR binding kinetic parameters (see Methods for details).

α	f_{H1}	f_{inter}	t_{sim} (ms)	k_+ (μs^{-1})	k_- (μs^{-1})	k_{cap} (μs^{-1})	k_{esc} (μs^{-1})	k_{evo} (μs^{-1})	N_{cap}	N_{evo}	ρ_{evo} (%)	$MFPT_{evo}$ (ns)
0.44	1.4	0.977	1.79	0.32 (0.02)	0.69 (0.04)	95.65 (0.41)	109.15 (0.55)	0.66 (0.04)	62484	366	0.56	28.66
0.49	1.5	0.973	1.76	0.32 (0.02)	0.60 (0.03)	94.15 (0.39)	110.35 (0.43)	0.67 (0.03)	58388	348	0.60	29.75
0.53	1.6	0.968	1.72	0.32 (0.02)	0.52 (0.03)	93.60 (0.44)	111.25 (0.49)	0.67 (0.04)	54762	321	0.59	28.72
0.69	2	0.959	1.70	0.36 (0.02)	0.39 (0.02)	93.10 (0.52)	111.25 (0.58)	0.78 (0.03)	44312	298	0.67	31.30
0.80	2.5	0.955	1.60	0.47 (0.04)	0.29 (0.01)	93.05 (0.70)	110.65 (0.68)	0.98 (0.08)	31481	264	0.84	31.07

t_{sim}: accumulated simulation time from 20 independent runs.
ρ_{evo}: probability of converting collision complex to fully bound state (i.e., N_{evo} / $N_{cap.}$)
Values in the parenthesis are standard errors of the mean, which is calculated as $\sigma/\sqrt{n}$, where σ is the standard deviation, and n is 20, i.e., the number of independent runs.

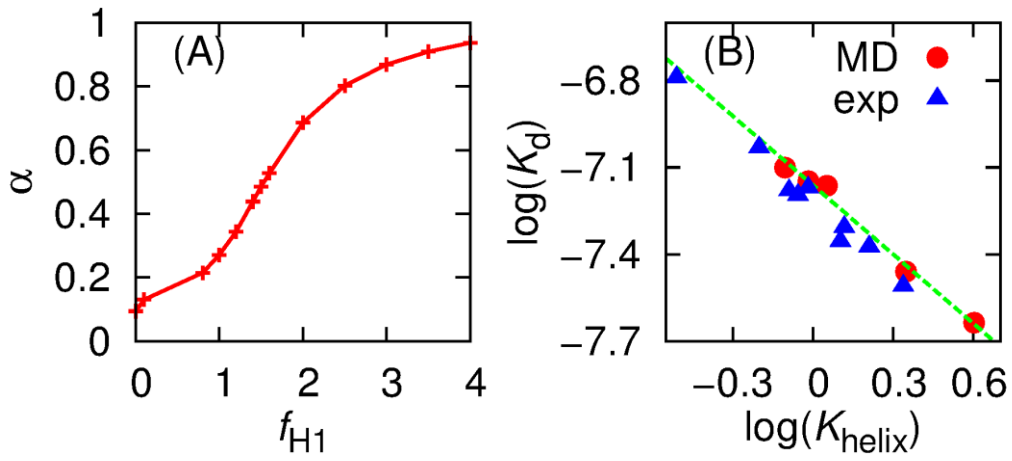


Figure 4.2 (A) H1 helicity of unbound ACTR (α) at 300 K versus the scaling factor f_{H1} of intra-molecular interaction strengths within ACTR H1 region. (B) log(K_{helix}) of H1 of unbound ACTR versus log(K_d) of ACTR/NCBD binding at 300 K from simulation and experiment[31]. The green line is the best-fitted line of simulation data. $K_{helix} = \alpha/(1-\alpha)$ and simulated K_d was calculated using Equation 4.1.

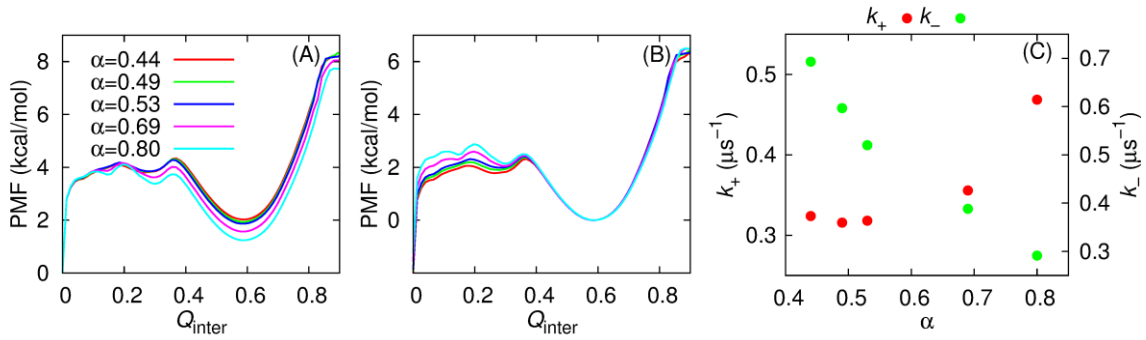


Figure 4.3 (A) Potential of mean force (PMF) as a function of Q_{inter} at 314 K for five NCBD/ACTR models. These profiles were obtained from MD production simulations and calculated as $-RT\ln[P(Q_{inter})]$, where $P(Q_{inter})$ is the probability distribution of Q_{inter}, T is the temperature, and R is the gas constant. All traces have been shifted such that the free energy value at Q_{inter}=0 is zero. (B) Same as (A) except that all traces have been shifted that free energy value at Q_{inter}=0.58 is zero. (C) Association and dissociation rates as a function of the H1 helicity of unbound ACTR.

4.2.2 Enhancing ACTR H1 helicity accelerates NCBD binding

An intriguing observation from the experimental study[31] is that stabilizing ACTR H1 increases the association rate constant, and at the same time decreases the dissociation rate constant of ACTR/NCBD complex formation[31]. Even though the Gō-like models were calibrated solely based on thermodynamic properties, they can reproduce the dependence of both association and dissociation rate on ACTR H1 helicity. As shown in Table 4.1 and Figure 4.3C, the association rate (k_+) is moderately enhanced as ACTR H1 becomes more helical, which is also consistent with decreasing binding free energy barriers as shown in Figure 4.3A. Free energy decomposition analysis (Figure 4.9) shows that the binding free energy barrier along Q_{inter} stemmed from the imperfect compensation of the favorable enthalpic and unfavorable entropic contributions, as predicted in the funnel-like free energy landscape theory of protein folding[227]. Stabilizing ACTR H1 results in more favorable interaction energies near the transition state (e.g., Figure 4.9A inset), thus

lowering the free energy barrier of association. The simulations also predict that the dissociation rate (k_-) decreases as α is increased from 0.44 to 0.80 (Figure 4.3C), which seems to be a direct consequence of increasing stability of the bound state (see Figure 4.3B).

4.2.3 Accelerated NCBD binding through efficient folding upon encounter

To understand how ACTR H1 helicity enhancement modulates the NCBD/ACTR binding kinetics, we further evaluate its effects on various stages of NCBD/ACTR complex formation. For this, three general states, including the unbound, collision complex and bound states, were defined (see Equation 4.3) and transitions between these states were extracted from the production trajectories and analyzed. The results are summarized in Table 4.1. For more disordered ACTR, the capture rate (k_{cap}) is slightly faster, and there were significantly more capture events (N_{cap}), which seems to be consistent with the fly-casting theory[24]. However, collision complex formation occurs considerably faster than forming the bound complex, and the correlation between capture rate and association rate is very week (Figure 4.10), implying that this step alone unlikely determines the overall kinetics of coupled binding and folding of ACTR to NCBD. On the contrary, the evolution rate (k_{evo}) is on the same order of magnitude as the overall binding rate (k_+) (Table 4.1 and Figure 4.4A) and shows the strongest correlation with k_+ (Figure 4.4A), suggesting that the impacts of increasing ACTR H1 helicity are mainly reflected in k_{evo}.

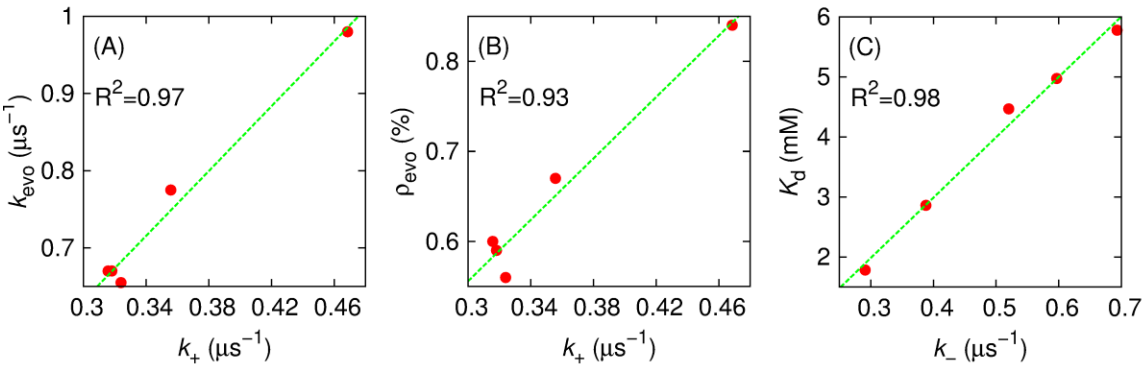


Figure 4.4 (A) Evolution rate and (B) probability of collision complex evolving into the bound state as a function of the associate rate. (C) Kd at 314 K versus the dissociation rate.

Note that both mean first passage time of evolution ($MFPT_{evo}$) and number of evolution events (N_{evo}) contribute to the calculation of k_{evo} (Equation 4.6), and we would like to further understand which one is the dominant factor that leads to increased k_{evo} (and k_+) induced by ACTR H1 stabilization. Intriguingly, $MFPT_{evo}$ is on the order of tens of nanoseconds in all cases (Table 4.1), suggesting that evolution from collision complex to bound state occurs rapidly for successful transitions. This agrees well with previous notion that many rare events are rare because they are infrequent, and not because they are slow (for instance, the protein folding transition path time was found to be 10,000 shorter than the mean waiting time in unfolded state) [228-230]. This observation also implies that NCBD/ACTR complex formation seems to be a sharp, cooperative structural transition, as suggested previously by Dogan and coworkers[231]. Therefore, $MFPT_{evo}$ cannot be used to explain the kinetic advantage of increased ACTR helicity on NCBD binding. As shown in Table 4.1, $MFPT_{evo}$ indeed doesn't decrease when ACTR is more helical, and similar insensitivity of $MFPT_{evo}$ in response to changes of ionic strength is also found in the case of PUMA binding to Mcl-1[232]. Instead, greater k_{evo} and k_+ arise

mainly due to the higher probability of converting collision complex to fully bound state (ρ_{evo}) for ACTR variants with more preformed H1 helical structures (Figure 4.4**B**). That is, ACTR residual structure accelerates NCBD binding mainly by promoting more efficient folding upon encounter.

The above findings from kinetic analysis are further supported by the thermodynamic and structural analysis. As shown in Figure 4.3A, the free energy profiles reveal the presence of a metastable intermediate state at $Q_{inter} \approx 0.3$, which appears to be critical for lowering binding free energy barriers as ACTR H1 helicity increases. As shown in Figure 4.5, when unbound ACTR has a higher level of secondary structures, more native contacts can be formed in both peptides at this intermediate state, which may explain the extra enthalpy gain (Figure 4.9A inset) that lowers the binding free energy barrier. Furthermore, the presence of energetically more favorable, structurally more ordered intermediate state may be more "folding-competent" both energetically and topologically, which is consistent with the higher probability of collision complex evolving into fully bound state induced by ACTR helicity enhancement (Table 4.1). Altogether, the current analysis reinforces the pivotal role of efficient folding upon encounter in accelerating the overall kinetics of coupled binding and folding[53, 209-211].

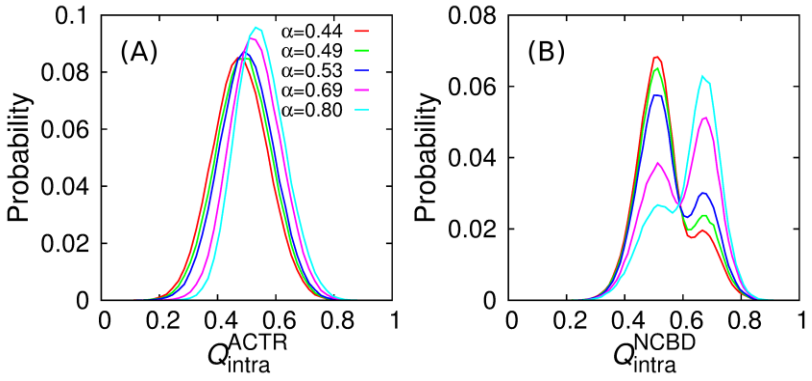


Figure 4.5 Distributions of the fraction of native contacts formed within ACTR (A) and NCBD (B) in the intermediated state ($0.21 \leq Q_{inter} \leq 0.32$).

The reduced dissociation rate as ACTR H1 becomes increasingly helical (Table 4.1) appeared to correlate well with the increased binding affinity between ACTR and NCBD (Figure 4.4C). This is not surprising considering that dissociation is largely a unimolecular process, where the rate is mainly determined by the depth of the bound free energy minimum[227]. In other words, reduction of dissociation rate is a direct consequence of stronger binding between NCBD and ACTR with increasing ACTR H1 helicity.

4.2.4 Baseline mechanism of NCBD/ACTR recognition

Although the topology-based models neglect many atomistic details and the absolute values presented here may not be directly compared with experimental results, these models were based on the principle of minimal frustration of protein folding[233-235] and should be able to capture the baseline mechanisms of coupled binding and folding of IDPs. The carefully calibrated Gō-like models seem to reproduce both thermodynamic and kinetic properties of NCBD/ACTR complex formation, thus allowing us to derive further mechanistic insights into NCBD/ACTR coupled binding and folding. As shown in Figure 4.6, the overall folding of each peptide is

coupled with the binding in a cooperative manner, as the fractions of total intra-molecular native contacts ($Q_{\text{intra}}^{\text{ACTR}}$ and $Q_{\text{intra}}^{\text{NCBD}}$) increase gradually with Q_{inter}. This is consistent with previous experimental[236] and simulation findings[81, 198, 210]. Such an overall mechanism is conserved as ACTR H1 becomes more helical (Figure 4.6). Moreover, detailed examination of the transition path ensemble suggests that the ensemble is heterogeneous and transitions at the microscopic level may involve multiple trials and steps (Figure 4.7). The presence of several intermediate states as well as involvement of both conformational selection (e.g., Figure 4.7A and B) and induced fit (e.g., Figure 4.7C and D) pathways seems to agree with the "extended conformational selection" model[237]. These mechanistic features have been described in detail in our previous work[198].

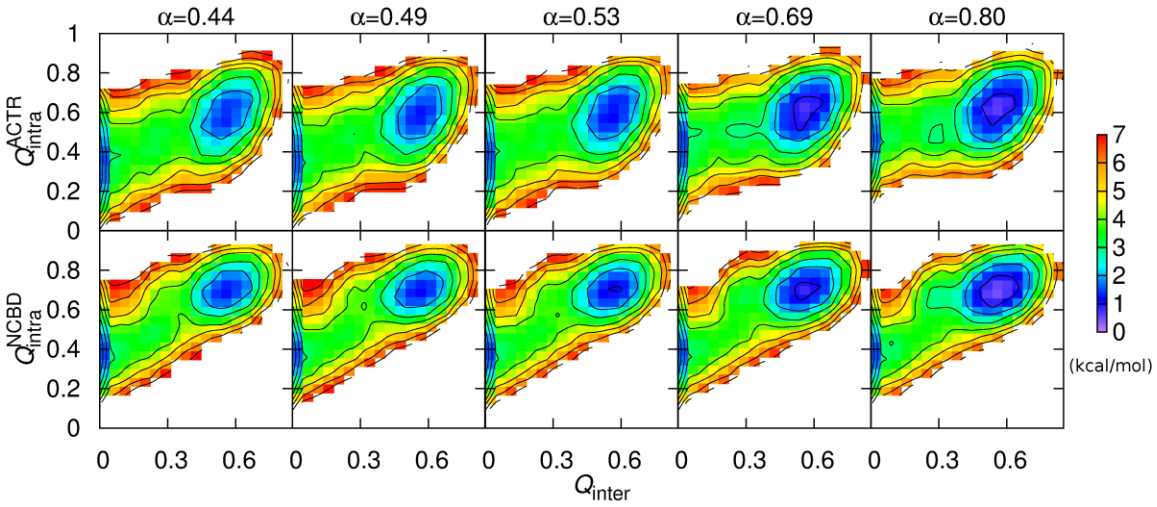


Figure 4.6 2D free energy profiles at 314 K as a function of Q_{inter} and $Q_{\text{intra}}^{\text{ACTR}}$ (top) or $Q_{\text{intra}}^{\text{NCBD}}$ (bottom). Contour levels are drawn at every 1 kcal/mol.

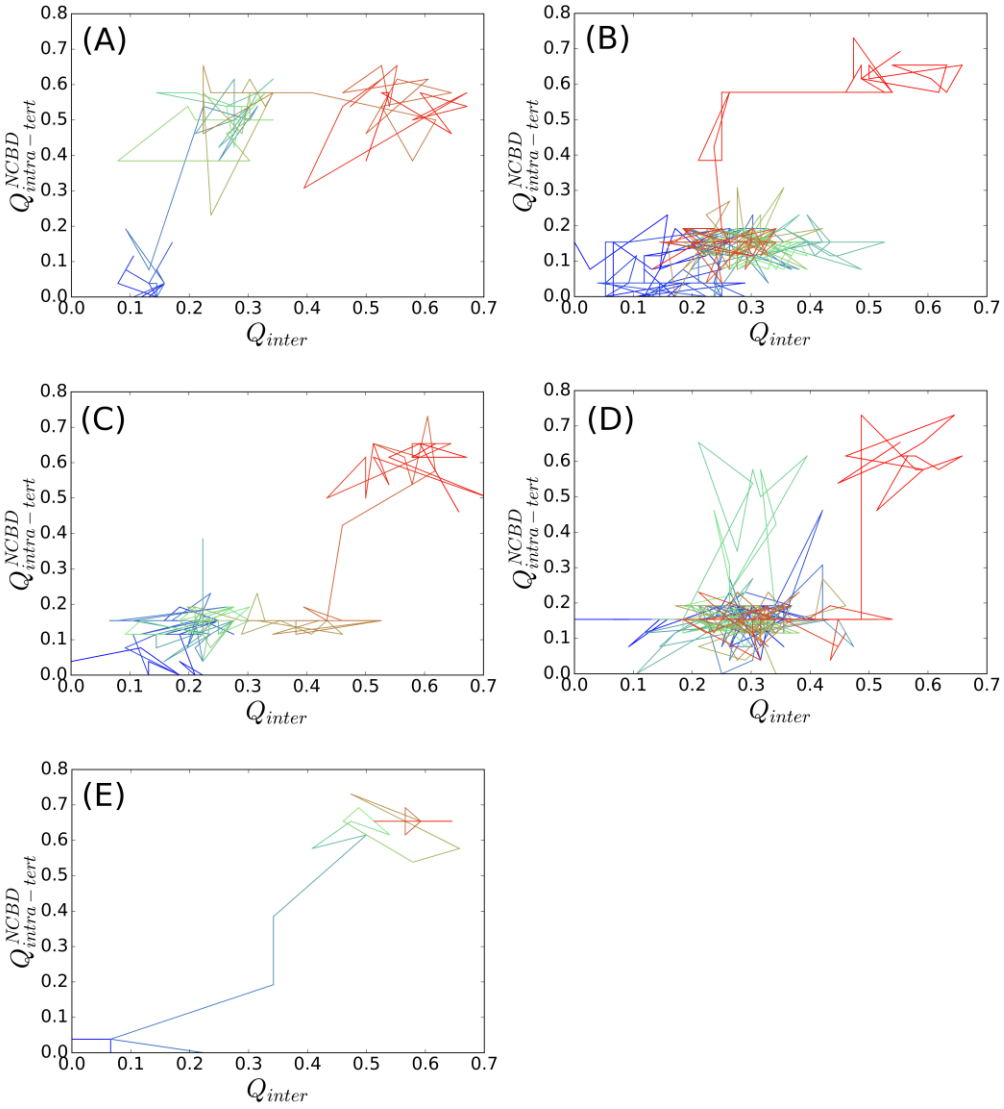


Figure 4.7 Representative transitions from unbound to bound state. The traces were color in time ordering from blue to red. (A) and (B) are for conformational selection like pathways, (C) and (D) for induced fit like pathways, and (E) for other pathways. Transient, but non-productive visit of alternative pathways was present in (B) and (D), but absent in (A) and (C).

Stabilizing ACTR H1 results in continued differences in the pathway of coupled binding and folding. We have analyzed all productive transitions from unbound to bound states by monitoring the change of NCBD intra-molecular tertiary structures in response to ACTR binding. Figure 4.8 shows the 2D free energy profiles as a

function of $Q_{\text{intra-tert}}^{\text{NCBD}}$ and Q_{inter}, which reveal how the relative populations of two key intermediates (I1 and I2 in Figure 4.8A) change with increasing ACTR H1 helicity. For more disordered ACTR H1, intermediate state I1 appears to be dominant, where most tertiary structures within NCBD were not formed (e.g., see Figure 4.8A). As ACTR H1 becomes increasingly helical, the population of intermediate state I2 increases, where NCBD has already gained significant amount of tertiary structures when only ~25% of native contacts between NCBD and ACTR have been formed (e.g., see Figure 4.8E). Such shift of relative populations of I1 and I2 also suggests that as ACTR H1 became more helical, there is a greater chance that NCBD could become folded before reach the fully bound state. This is consistent with the previous notion that enhancing the helicity of ACTR H1 results in an intermediate state with "folding-competent" topology, thus promoting efficient folding upon encounter and accelerating its association with NCBD.

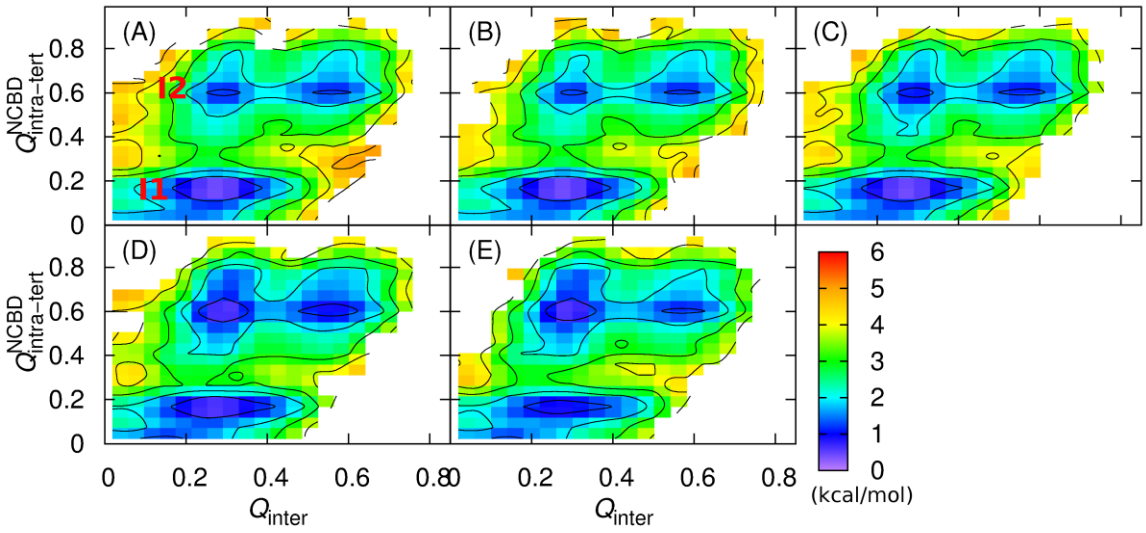


Figure 4.8 2D free energy profiles at 314 K as a function of $Q_{\text{intra-tert}}^{\text{NCBD}}$ and Q_{inter} calculated from transition path ensembles. Panels A-E are for α = 0.44, 0.49, 0.53, 0.69 and 0.80, respectively. Contour levels are drawn at every 1 kcal/mol.

4.3 Conclusions

Recent NMR structural studies and stopped-flow kinetic measurements have shown that increasing helicity of ACTR H1 not only enhances the stability of NCBD/ACTR complex, but also accelerates the association rate while decelerating the dissociation rate[31]. This is surprising because previous studies have generally emphasized the importance of conformational flexibility in IDP recognition and stabilizing residual helicities have been found to either reduce the binding kinetics or have minimal impacts[70, 203, 208]. Here we have constructed a series of topology-based coarse-grained models that were carefully calibrated by reproducing key thermodynamic properties of the unbound state and NCBD/ACTR interaction. Through milliseconds of MD simulations, hundreds of reversible binding and folding transitions were generated to analyze the kinetics, thermodynamics and mechanism of NCBD/ACTR interaction. The results show that with an increasing amount of preformed structural elements in unbound IDPs, the overall entropic cost of forming NCBD/ACTR complex is reduced, which leads to increased binding affinity and thus reduced dissociation rate. Increasing ACTR helicity does not significantly alter the baseline mechanism of the synergistic folding of ACTR and NCBD during association, which continues to follow an “extended conformational selection” model with multiple stages of selection and induced folding. Increasing residual structure in ACTR H1, however, results in a higher probability of productive evolution of nonspecific collision complexes to the final bound and folded state, and this is shown to be the primary source of accelerated association kinetics. Taken together, this study provides mechanistic insights into how residual structure modulates

thermodynamics and kinetics of coupled binding and folding of IDPs and highlights the importance of efficient folding upon encounter in such processes.

4.4 Methods

4.4.1 Topology-based coarse-grained models of NCBD/ACTR

A Cα-only sequence-flavored Gō-like model was previously derived for the wild-type NCBD/ACTR complex (see Figure 4.1)[198]. Briefly, this model was built based on PDB structure 1KBH[218] (model 1) using the MMTSB Gō-Model Builder (http://www.mmtsb.org)[235, 238]. It was then calibrated to recapitulate the overall residual structure level of unbound peptides and the binding affinity. This was achieved by first uniformly scaling the interaction strengths of intra-molecular native contacts, to reproduce the experimental residual helicity profiles of unbound peptides. The interaction strengths of inter-molecular native contacts were then scaled to match the simulated binding affinity of the complex with experimental values. In this model, the total numbers of native contacts between peptides (N_{inter}), within NCBD ($N_{\mathrm{intra}}^{\mathrm{NCBD}}$), within ACTR ($N_{\mathrm{intra}}^{\mathrm{ACTR}}$) and in NCBD intra-molecular tertiary structures ($N_{\mathrm{intra-tert}}^{\mathrm{NCBD}}$) are 76, 78, 49 and 26, respectively.

To model ACTR variants with different H1 residual helicities, the above Gō-like model was further tuned by uniformly scaling its interaction strengths of all intra-molecular native contacts within segment H1. Scaling factors for ACTR H1 (f_{H1}) were set at 1.4, 1.5, 1.6, 2.0 and 2.5 for five ACTR variants, which yield H1 residual helicities (α) ranging from 0.44 to 0.80 (see Table 4.1), comparable to the experimental values for mutants studied by Kjaergaard and coworkers[31]. For each model, the inter-molecular interaction strengths between ACTR variant and NCBD

were uniformly rescaled, such that the calculated thermodynamic stability of the NCBD/ACTR complex matched the experimental value[31]. This fine-tuning was performed using replica exchange (REX) simulations in combination with Hamiltonian mapping[226, 239]. The inter-molecular scaling factors (f_{inter}) in the five final models are 0.977, 0.973, 0.968, 0.959 and 0.955, respectively (see Table 4.1).

4.4.2 Simulation protocols

All MD simulations were performed using CHARMM[149, 150], and REX simulations were performed using CHARMM[149, 150] with MMTSB[146]. Langevin dynamics was used with a friction coefficient of 0.1 ps^{-1} and a time step of 10 fs. All bond lengths were constrained using SHAKE[156].

To compute the averaged helicities of ACTR H1, 1 μs MD simulations of ACTR alone were performed at 300 K. For all simulations of NCBD/ACTR, the complex was put in a cubic box with the size of 105 Å under periodic boundary conditions. First, a REX simulation of the complex was performed for 2 μs using the original Gō-like model[198] to generate structural ensembles. Hamiltonian mapping[226, 239] was then used to identify scaling parameters that would reproduce the experimental stabilities. 2 μs REX simulations using these new models were then carried out to verify the stability of the complex. After model calibration, all production simulations of the NCBD/ACTR complexes were performed at 314 K, near the melting temperatures. For each model, 20 independent simulations were initiated from random conformations. The accumulated simulation times ranged from 1.60 to 1.79 ms for each complex (Table 4.1), which yielded hundreds of

reversible binding/unbinding transitions for reliable analysis of kinetic rates and pathways. Simulation trajectories were saved every 100 ps for analysis.

4.4.3 Analysis

To calculate the average helicity, we first calculated the number of (i, i+4) Cα-Cα contacts in ACTR H1. A contact is considered to be formed if the Cα-Cα distance is no more than 1 Å greater than that in the fully folded state (PDB: 1KBH). The overall helicity of ACTR H1 was then calculated as the ensemble-averaged fraction of (i, i+4) Cα-Cα native contacts formed.

For the complex, the dissociation constants (K_d in M) were computed from the bound probabilities (p_b) as[198]

$$K_d = \frac{1660}{V_0}\frac{(1-p_b)^2}{p_b} \tag{4.1}$$

where V_0 is the volume of simulation box in Å^3. Conformations with $N_{\text{inter}} \geq 44$ are considered bound.

All kinetic information was derived from production simulations of the complex. Both the binding (k_+) and unbinding rates (k_-) were calculated directly as the inverse of corresponding mean first passage times (MFPTs) between the bound (B) and unbound (U) states (Equation 4.2), where the two states were defined as $N_{\text{inter}} \geq 44$ and $N_{\text{inter}} < 1$, respectively. Running average over 10 ns was performed before assigning the states to suppress fictitious high frequency transitions.

$$U \overset{k_+/k_-}{\longleftrightarrow} B \tag{4.2}$$

To further understand how residual helical stability may affect different stages of coupled binding and folding, three general conformational states were defined, including an additional collision complex (CC) state,

$$U \overset{k_{cap}/k_{esc}}{\longleftrightarrow} CC \overset{k_{evo}}{\longrightarrow} B \tag{4.3}$$

where k_{cap}, k_{esc}, and k_{evo} are capture, escape and evolution rates, respectively. Here, the unbound state was more strictly defined as conformations without specific or nonspecific inter-molecular contacts (i.e., $N_{\text{inter}} < 1$ and $N_{\text{inter}}^{\text{ns}} < 1$). CC includes conformations with only nonspecific inter-molecular contacts (i.e., $N_{\text{inter}} < 1$ and $N_{\text{inter}}^{\text{ns}} \geq 1$). Again, running average over 10 ns was performed before state assignment. k_{cap}, k_{esc}, and k_{evo} were calculated from number of transitions and MFPT between states as described previously[210, 220]:

$$k_{\text{cap}} = \text{MFPT}_{\text{cap}}^{-1} \tag{4.4}$$

$$k_{\text{esc}} = \frac{N_{\text{esc}}/(N_{\text{esc}} + N_{\text{evo}})}{(\text{MFPT}_{\text{esc}} \times N_{\text{esc}} + \text{MFPT}_{\text{evo}} \times N_{\text{evo}})/(N_{\text{esc}} + N_{\text{evo}})} \tag{4.5}$$

$$k_{\text{evo}} = \frac{N_{\text{evo}}/(N_{\text{esc}} + N_{\text{evo}})}{(\text{MFPT}_{\text{esc}} \times N_{\text{esc}} + \text{MFPT}_{\text{evo}} \times N_{\text{evo}})/(N_{\text{esc}} + N_{\text{evo}})} \tag{4.6}$$

N_{cap}, N_{esc}, and N_{evo} are the numbers of capture, escape, and evolution transitions, respectively.

Ensembles of transition path trajectories were collected from production MD simulations to further analyze the pathways of NCBD/ACTR synergistic folding.

Each transition path trajectory was defined between the fully unbound state (N_{inter} = 0 and $N_{\mathrm{inter}}^{\mathrm{ns}}$ = 0) last visited and the fully bound state ($N_{\mathrm{inter}} \geq 50$) first visited (e.g., see Figure 4.7). The more stringent state criteria are necessary to eliminate spurious, noncommitting transitions. Very few intra-molecular tertiary contacts were present in bound state of ACTR[198], thus we didn't examine ACTR folding upon binding. NCBD is a molten globular protein with a high level of secondary structures in the unbound state. Its binding-induced folding was analyzed by tracking its intra-molecular tertiary structure formation, quantified by $Q_{\mathrm{intra-tert}}^{\mathrm{NCBD}}$.

4.5 Supporting Material

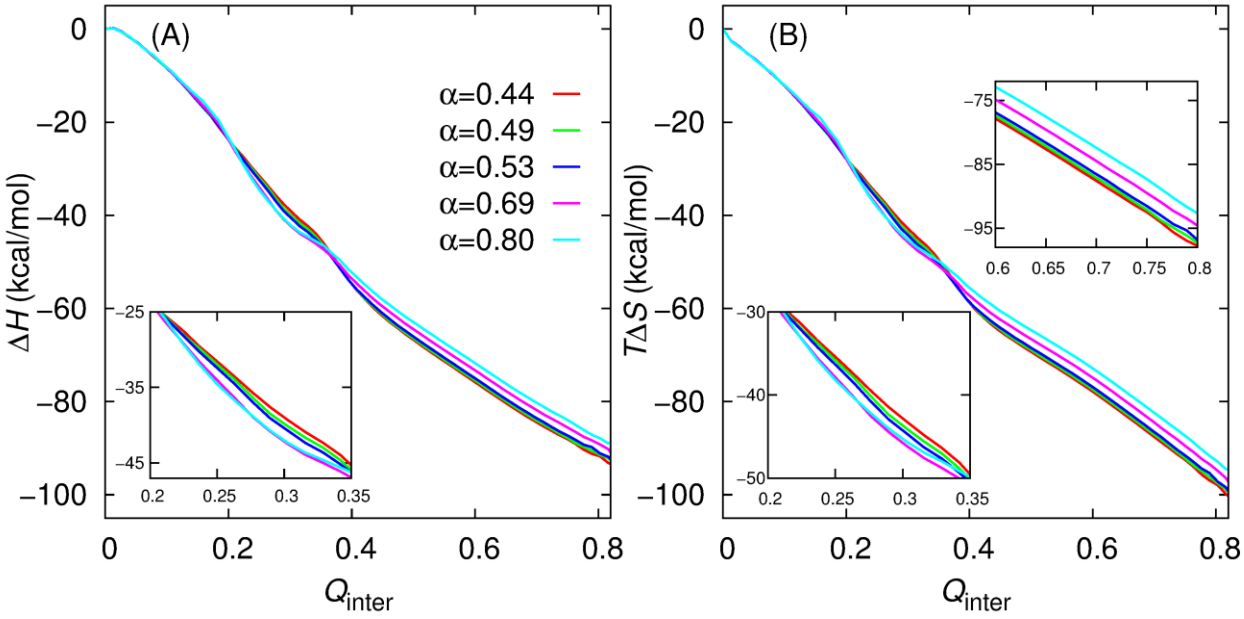


Figure 4.9 Enthalpic (A) and entropic (B) contributions in NCBD/ACTR complex formation. Here, we used the unbound state (Q_{inter} = 0) as the reference state. Enthalpic contribution (ΔH) was obtained from the averaged potential energies for each Q_{inter} state. Entropic contribution ($T\Delta S$) was calculated as the difference between ΔH and free energy values.

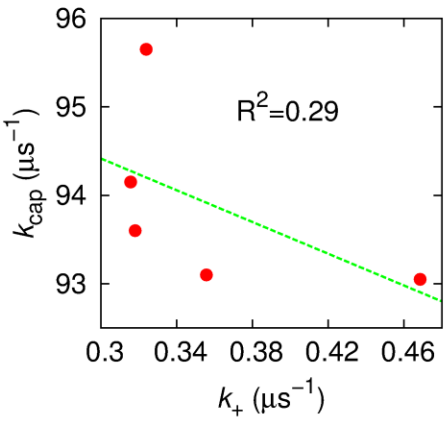


Figure 4.10 Capture rate as a function of the associate rate.

CHAPTER 5

HYRES: A COARSE-GRAINED MODEL FOR MULTI-SCALE ENHANCED SAMPLING OF DISORDERED PROTEIN CONFORMATIONS*

Efficient coarse-grained (CG) models can be coupled with atomistic force fields to accelerate the sampling of atomistic energy landscapes in the multi-scale enhanced sampling (MSES) framework. This approach may be particularly suitable for generating atomistic conformational ensembles of intrinsically disordered proteins (IDPs). While MSES is relatively robust to inherent CG artifacts, achieving optimal sampling efficiency requires CG modeling to generate local and long-range fluctuations that are largely consistent with those at the atomistic level. Here, we describe a new hybrid resolution CG model (HyRes) for MSES simulations of disordered protein states, which is specifically designed to provide semi-quantitative secondary structure propensities together with a qualitative description of long-range nonspecific interactions. The HyRes model contains an atomistic description of the backbone with intermediate resolution side chains. The secondary structure propensities are tuned by adjusting the backbone hydrogen-bonding strength and ϕ/ψ torsion profile. The sizes and covalent geometries of side chains are parameterized to reproduce distributions derived from atomistic simulations. Lennard-Jones parameters for sidechain beads are assigned to reproduce statistical potentials derived from protein structural database, and then globally parameterized with

* Liu X. and Chen J (2017). "HyRes: A Coarse-grained Model for Multi-scale Enhanced Sampling of Disordered Protein Conformations." Physical Chemistry Chemical Physics 19(48): 32421-32432.

nonspecific electrostatic interactions to reproduce free energy profiles of pair wise interactions and key conformational properties of model peptides. Application of HyRes to MSES simulations of small IDPs suggests that it is capable of driving faster structural transitions at the atomistic level and increasing the convergence rate compared to Cα-only Gō-like models previously utilized. With further optimization, we believe that the new CG model could greatly improve the efficiency of MSES simulations of larger and more complex IDPs frequently involved in cellular signaling and regulation.

5.1 Introduction

Intrinsically disordered proteins (IDPs) remain fully or partially disordered in the unbound state under physiological conditions[2, 3, 21, 53]_ENREF_1. Compared with globular proteins, IDPs are enriched in polar and charged residues, and deficient in hydrophobic residues[11]. They often play significant roles in cellular signalling and regulation[17, 21, 48, 107, 170, 175], and are frequently associated with human diseases including cancer and neurodegenerative diseases[13, 170, 173, 174, 240]. There is thus a great need to understand the physical basis of IDP functions in molecular recognition and cellular signaling. The dynamic nature of unbound IDPs, however, requires their properties to be described by heterogeneous ensembles that are not amenable to traditional experimental techniques geared towards structured proteins. Averaged properties measured on disordered protein states are generally insufficient to uniquely define the underlying ensembles via various restrained structure calculation or selection selection protocols[7, 53, 55,

206, 241] ENREF 15. Reliable physics-based atomistic simulations thus have an important role to play.

We recently adapted the multi-scale enhanced sampling (MSES) approach of Kidera and coworkers[242] to couple Gō-like coarse-grained (CG) models for enhanced sampling of protein conformations in atomistic (AT) force fields[83, 243]. In brief, both AT and CG representations of the system are simulated simultaneously, but they don't interact directly with each other except through a MSES coupling potential (U_{MSES}). Potential energy of the hybrid system is shown in Eq. 5.1,

$$U_{hybrid}(r_{AT}, r_{CG}, \lambda) = U_{AT}(r_{AT}) + U_{CG}(r_{CG}) + \lambda U_{MSES}(r_{AT}, r_{CG}) \quad (5.1)$$

where UAT and UCG are AT and CG potentials, respectively. UMSES is designed to restrict the structural divergence between AT and CG models along selected degrees of freedom. A proper UMSES is expected to leverage rapid CG fluctuation to accelerate structural transitions at the atomistic level, and the scaling factor λ determines how closely the AT and CG configurations could track each other. Bias introduced by MSES coupling is removed using Hamiltonian replica exchange (H-REX), which communicates coupled conformational dynamics to the limit of λ=0 where AT and CG models are completely independent and proper ensembles could be recovered at both resolutions. Our MSES approach also incorporates temperature replica exchange (T-REX) to further enhance the sampling efficiency[83, 243]. Application to small β-hairpins and helical peptides supports that MSES could dramatically accelerate the sampling of peptide conformational transitions. Compared with other multiscale sampling methods such as direct coordinate swapping in resolution exchange[244-246], coupling AT and CG models through

U_{MSES} provides several significant advantages. It offers flexibility in controlling energetic impact of diverged conformations at AT and CG levels, which improves exchange efficiency and provides superior scalability to large systems. Also, it prevents the CG model from dictating the conformational dynamics and makes MSES more robust against inherent CG artifacts.

A key bottleneck in sampling large-scale conformational rearrangement, either for folded proteins or IDPs, is that these transitions are often entropically limited[83, 243, 244]. Kinetics of such transition is temperature insensitive, and the popular temperature-based REX protocols becomes ineffective. In this regard, MSES could provide unique advantages, because CG modeling reduces the conformational space and may be particularly effective in driving conformational search at atomistic level. However, the efficiency of MSES depends critically on the ability of CG modeling to generate local and long-range structural fluctuations that are largely consistent with those expected at the atomistic level. Otherwise, the CG model would attempt to drive atomistic transitions through unfavorable pathways, which will hinder the convergence of MSES simulations. Previous studies have shown that coupling the AT model to a simple Cα-based Gō-like model could significantly increase atomistic structural transitions in simulations of both β-hairpin[83] and small IDPs[243]. However, these studies also suggest that the efficiency of MSES for simulating IDPs is limited by the existence of inconsistent structural fluctuations between Cα-based CG and atomistic models[243]. This problem is anticipated to become more severe for larger and more complex IDPs. There is thus a need to explore or develop alternative CG models for more efficient MSES simulations of IDPs.

CG modeling is an area of intensive research and numerous CG protein models have been described in the literature. Some CG force fields treat solvent molecules implicitly, for example, the united-residue (UNRES) model[247-252], PRIME (protein intermediate resolution model)[253-256], Bereau and Deserno's CG model[257], OPEP (optimized potential for efficient structure prediction) force field[258-262] and PRIMO force field[263]. Others include solvent molecules explicitly, such as MARTINI[264, 265] and PACE[266]. We refer the reader to references[267-270] for more complete reviews. We note that virtually all existing CG models are designed to work independently (i.e., as alternatives to AT models), which presents formidable challenges and requires stringent optimization and/or more sophisticated effective potential energy functions. We have evaluated several CG protein models and the results suggest that none of them is capable of provide a balanced description of disordered protein ensembles[271]. Here, we describe a new hybrid resolution CG model (HyRes) that has been specifically designed for MSES simulations of IDPs. A guiding hypothesis is that, the CG model should be able to capture essential conformational features of IDPs to achieve best possible efficiency of MSES simulations. In particular, most regulatory IDPs are known to contain some level of residual secondary structures (mostly helical) and transient long-range contacts with functional implications[2, 7, 81, 272]. Importantly, the new model does not have to be highly accurate to be effective in MSES simulations, because the quality of structure ensembles derived from MSES simulations is ultimately determined by the AT model. In addition, HyRes should be designed with simple potential energy function to minimize the computational overhead in MSES

simulations. To this end, the new model contains a united-atom representation of the backbone to semi-quantitatively describe secondary structure propensities. Side chains are represented at intermediate resolution, to provide a qualitative description of long-range nonspecific interactions. The model is parameterized based on atomistic and CG simulations of model compounds and peptides as well as statistical potentials derived from protein structure database. The new model is then validated using a set of model IDPs and its efficacy for MSES simulations evaluated using short IDPs.

5.2 Models and Methods

5.2.1 Peptide model with hybrid resolutions

In the HyRes model, the peptide backbone is represented at the united-atom level and side chains at an intermediate resolution. The goal is to allow semi-quantitative description of secondary structure propensities together with a qualitative description of geometries and pair-wise interactions between side chains. Detailed description of peptide backbone, particularly the hydrogen bonding geometry, has been shown to be critical for realistic modeling of peptide conformational equilibrium[273]. Figure 5.1 summarizes the representation of all 20 amino acids in HyRes. There is no side chain bead in Gly, while there are two side chain beads in Lys and Arg, three side chain beads in His, Phe, and Tyr, five side chain beads in Trp, and one side chain bead in the other amino acids. Detailed mapping of side chain atoms from the atomistic model[152, 153, 274] to CG beads are summarized in Table 5.6.

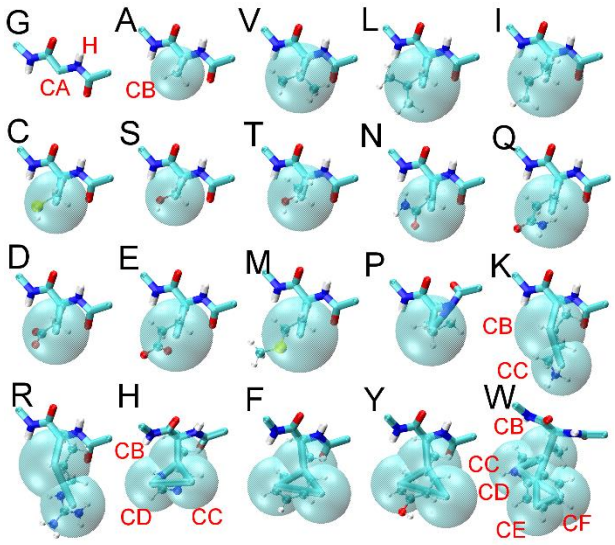


Figure 5.1 HyRes representation of all 20 natural amino acids. The CG side chain beads are shown in VDW representation to proportion, with the atomistic model shown in ball-and-stick to illustrate the mapping scheme (summarized in Table 5.6). Representative backbone and side chain beads in the HyRes model are labeled in red.

5.2.2 Effective potential energy function

The total interaction energy U is shown in Eq. 5.2. It consists of eight terms, including bond (U_{bond}), angle (U_{angle}), dihedral (U_{dihedral}), improper dihedral (U_{improper}), backbone ϕ/ψ torsion cross-term (U_{CMAP}), Lennard-Jones interactions (U_{LJ}), hydrogen bonding (U_{Hbond}) and electrostatic potentials (U_{elec}),

$$U = U_{bond} + U_{angle} + U_{dihedral} + U_{improper} + U_{CMAP} + U_{Hbond} + U_{LJ} + U_{elec} \quad (5.2)$$

$$U_{bond} = \sum_{bonds} k_b (b - b_0)^2 \quad (5.3)$$

$$U_{angle} = \sum_{angles} k_\theta (\theta - \theta_0)^2 \quad (5.4)$$

$$U_{dihedral} = \sum_{dihedrals} k_\chi [1 + cos(n\chi - \delta)] \quad (5.5)$$

$$U_{improper} = \sum_{impropers} k_\psi (\psi - \psi_0)^2 \quad (5.6)$$

$$U_{Hbond} = \sum_{Hbonds} \varepsilon_{HB} \left[5 \left(\frac{\sigma}{r} \right)^{12} - 6 \left(\frac{\sigma}{r} \right)^{10} \right] cos^4 \theta_{AHD} \quad (5.7)$$

$$U_{\mathrm{CMAP}} = \sum_{\substack{\text{non-Gly, non-Pro} \\ \text{residues}}} U_{\mathrm{CMAP}}(\varphi, \psi) \quad (5.8)$$

The bond (U_{bond}) and angle (U_{angle}) interactions are described as harmonic potentials (Eq. 5.3 and 5.4). U_{angle} is found between all adjacent bonds, except those constituting a triangle (such as for the His side chain) and the C_{i-1}-N_i-CB_i angle for Pro. Bond and angle parameters for the peptide backbone are obtained directly from the CHARMM19 united-atom force field[275]. The equilibrium values and force constants of other bonds and angles are optimized to reproduce distributions obtained from atomistic simulations of dipeptides in the GBSW implicit solvent[59] based on the mapping scheme (Table 5.6). The force constants are restricted to be between 200 to 600 kcal/mol/Å^2 for bonds and to 20-90 kcal/mol/rad^2 for angles during the optimization.

Dihedral potentials (U_{dihedral} shown in Eq. 5.5) are applied to peptide ϕ (C_{i-1}-N_i-CA_i-C_i), ω (CA_{i-1}-C_{i-1}-N_i-CA_i), χ (N_i-CA_i-CB_i-CC_i, if present) and ψ' (CB_i-CA_i-C_i-O_i,

only for a few selected residues, see below). Parameters related to ϕ are obtained directly from the CHARMM 19 force field[275]. The peptide bond strongly favors the *trans*-conformation for all non-proline amino acids. Therefore, for ω, n is set to 2 and δ to 180° in Eq. 5.5, and a 9.00 kcal/mol energy barrier is used here to separate *cis*- and *trans*- conformations. The barrier reflects a slight increase from the 8.2 kcal/mol value in CHARMM19, such that only *trans*-conformation is sampled in CG simulations of dipeptides. However, for ω in X_{i-1}-Pro_i, both *cis*- and *trans*-conformations are accessible. To allow efficient sampling of both *cis*- and *trans*-conformations, a smaller energy barrier (5.40 kcal/mol) is applied to ω in this case, which is based on the previous report that the barrier of *trans*- to *cis*- rotation for X_{i-1}-Pro_i is ~65% of that for X_{i-1}-$nonPro_i$[276]. U_{dihedral} related to χ is applied to residues Lys, Arg, His, Phe, Tyr and Trp, and parameters were obtained by fitting the distribution of χ from atomistic simulations of dipeptides.

Improper potentials (U_{improper}) are required to maintain the chirality of Cα atoms in non-Gly amino acids (i.e., on CA_i-N_i-C_i-CB_i, where CA_i is the central atom), planarity of peptide bonds (i.e., on N_i-C_{i-1}-CA_i-H_i for non-Pro amino acids and C_i-CA_i-N_{i+1}-O_i for all amino acids), and planarity of the Trp side chain ring. Although representing multiple atoms in the side chain by one bead could impact the geometry of Cα-connected atoms/beads to different degrees for different amino acids, we treat the chirality of Cα atoms uniformly (i.e., 35.3°) for all non-Gly amino acids in the HyRes model.

The secondary structure propensities are mainly determined by the backbone hydrogen bonding interactions (U_{Hbond}) and ϕ/ψ torsion cross-term (via the CMAP

term in CHARMM[153]) (U_{CMAP}). U_{Hbond} is applied on peptide backbone N, H and O atoms, and depends on both the N-O distance and O-H-N angle (r and θ_{AHD} in Eq. 5.7, respectively). σ in Eq. 5.7 is set to the ideal hydrogen bond donor-acceptor distance of 2.9 Å. In the current model, the same CMAP is applied to all non-Gly, non-Pro residues. The main role of the CMAP term is to fine tune secondary structure preferences together with backbone hydrogen-bonding interactions. Specifically, stabilizing or dis-stabilizing potentials may be added to ϕ/ψ regions corresponding to certain secondary structures, such as α-helix, π-helix, parallel and anti-parallel β strand, and poly-proline II (PPII) structures.

Nonspecific van der Waals (vdW) interactions between atoms separated by at least 3 virtual bonds are described by Lennard-Jones (LJ) potential (U_{LJ}) (Eq. 5.9), where $r_{\mathrm{min}}^{\mathrm{i}}/2$ and ε_{i} are the vdW radius and interaction strength of bead i. LJ parameters of the backbone atoms were first obtained from the CHARMM 19 force field, and then the interaction strengths were scaled down to account for lack of solvent screening of protein-protein dispersion interactions (see **Results and Discussions** for details). Initial vdW radii of side chain beads are assigned by matching the total volume occupied by corresponding atoms in the AT model, and ε_{i} values assigned based on Miyazawa and Jernigan's (MJ) statistical potentials[277]. These parameters are then optimized based on free energy profiles of pair-wise interactions and model peptide simulations (see **Results and Discussions** for details).

$$U_{LJ} = \sum_{i,j} \varepsilon_{i,j} \left[\left(\frac{r_{i,j}^{min}}{r_{i,j}} \right)^{12} - 2 \left(\frac{r_{i,j}^{min}}{r_{i,j}} \right)^{6} \right]$$

$$with\ r_{i,j}^{min} = \frac{r_i^{min} + r_j^{min}}{2}, and\ \varepsilon_{i,j} = \sqrt{\varepsilon_i \varepsilon_j} \quad (5.9)$$

There is no charge on either backbone atoms or side chain CG beads, except for those corresponding to charged groups in Lys, Arg, Asp, and Glu. Nonspecific electrostatic interactions are described using the Debye-Hückel-type potential,

$$U_{elec} = \sum_{chrg.pairs} \frac{q_i q_j}{4\pi\varepsilon_r \varepsilon_0 r_{i,j}} e^{-\frac{r_{i,j}}{D}} \quad (5.10)$$

where q_{i} is the charge of CG bead i, $r_{\mathrm{i,j}}$ is the inter-charge distance, and ε_0 is the permittivity of vacuum. *D* is the Debye screening length. At 300 K, *D* (in Å) depends ionic strength (*I*, in M) according to $\sqrt{9.48/I}$. Note that similar potentials have been widely used in various CG models and shown to provide an adequate description of long-range electrostatic interactions, even though accurate modeling of electrostatic solvation in heterogeneous protein environments requires more detailed and computationally more intensive implicit solvent theories. In the present work, *I* is set to 0.01 M, to roughly account for the low salt conditions in many NMR experiments.ε_{r} is the effective dielectric constant. The final value of ε_{r}, 20, was chosen in the current model based on simulations of several highly charged model peptides (see **Results and Discussions** for details).

5.2.3 AT and CG simulation protocols

The HyRes model has been implemented in CHARMM[149, 150]. Multiple sets of AT and CG simulations were performed in explicit or implicit solvent for backbone

and side chain analogs, dipeptides, and various model peptides. The protocols of these simulations are summarized below. Unless otherwise noted, Langevin dynamics was always used with a friction coefficient of 0.2 ps^{-1} and a time step of 2 fs. SHAKE algorithm[156] was used to constrain the length of all bonds involving hydrogen atoms.

• *Dipeptide simulations:* Explicit solvent simulations of the Ala dipeptide (Ace-Ala-Nme) were performed using both AMBER ff99SB-ILDN[74] and CHARMM36m[73] force fields, as implemented in GROMACS 2016.1[278]. 586 TIP3P water molecules[279] were used to solvate the dipeptide. Simulations were performed in the NPT ensemble at 298 K and 1 atm pressure under periodic boundary conditions. The PME method[280] was used to treat long-range electrostatic interactions. Short-range nonbonded interactions were truncated at 1.0 nm in the AMBER ff99SB-ILDN simulation, and 1.2 nm in CHARMM36m simulation. Each simulation lasted for 200 ns. In addition, all 20 dipeptides were simulated at 300 K for 300 ns in the GBSW implicit solvent, to derive the distributions of side chain conformations for parameterization of bonding parameters as well as for initial assignment of vdW radii of side chain beads.

• *Free energy of side chain interactions:* Potentials of mean force (PMFs) for pair-wise interactions between hydrophobic side chains were calculated by umbrella sampling simulations at 300 K in both the CHARMM36 explicit solvent and GBSW implicit solvent[59] atomistic force fields, and compared to the results from the HyRes model. In atomistic simulations, harmonic potentials with a force constant of 5.0 kcal/mol/$Å^2$ were imposed on the center of mass distance between the pair of

side chain analogs. 17 windows were used, spanning from 4 to 12 Å at a spacing of 0.5 Å. The simulation time at each window was 5 and 8 ns for explicit and implicit solvent calculations, respectively. In the CG simulation, the force constant of harmonic restraining potentials was 1.0 kcal/mol/Å^2. A total of 12 windows were used, spanning the reaction coordinate from 4 to 15 Å at a spacing of 1 Å. Each window was sampled for 200 ns. The weighted histogram analysis method (WHAM)[281] was then utilized to calculate the free energy profiles along the center of mass separation distance.

• *Model peptide simulations:* Multiple model peptides were used in parameterization and validation of HyRes, including Gly_{10}, $(AAQAA)_3$, 30 variants of the $(EK)_{25}$ peptide[64], the 24-residue arginine/serine (RS) peptide, the 28-residue kinase inducible transactivation domain of transcription factor CREB (KID), the activation domain of the activator for thyroid hormone and retinoid receptors (ACTR, residues 1040-1086), nuclear coactivator binding domain of the CREB binding protein (NCBD, residues 2059-2117), and a β-hairpin GB1m3[282]. The sequences of all model peptides are summarized in Table 5.7. Conformational ensembles of relatively shorter and simpler Gly_{10}, $(AAQAA)_3$, RS peptide, KID, and GB1m3 peptides in HyRes were calculated using direct MD simulations at 300 K. For each peptide, two independent simulations were performed, one starting from the folded or helical structure (control run) and the other starting from fully extended structure (folding run). The folded configuration of KID was obtained from RCSB Protein Data Bank (PDB) entry 1KDX[283]. Comparison of results from folding and control runs demonstrate that convergence can be achieved with 200 ns MD.

Conformational ensembles of ACTR, NCBD and $(EK)_{25}$ peptides in HyRes were calculated using T-REX simulations, performed using CHARMM in combination with the MMTSB toolset[146]. For $(EK)_{25}$ peptides, T-REX simulations were initiated from fully extended structures. For ACTR and NCBD, both control and folding runs were performed for convergence analysis. The folded conformations in control runs of ACTR and NCBD were obtained from their complex structure (PDB: 1KBH[284]). Each simulation involved eight replicas with temperatures exponentially spaced between 300 and 500 K, and exchanges between neighboring replicas were attempted every 2 ps. The simulations lasted for 200 ns per replica. Averaged exchange acceptance ratio was ~20% for all three peptides. Conformations sampled at the lowest temperature were collected for analysis.

For Gly_{10}, four additional 10 ns explicit solvent simulations were performed at 300 K under NPT conditions using the CHARMM36m force field to estimate solvent screening of dispersion interactions. The peptide was solvated by 7240 TIP3P water molecules in a 60 Å-cubic box. In these simulations, the peptide end-to-end distance was restrained to either 7 or 20 Å to represent the compact or extended conformation states. We also carried out Monte Carlo (MC) simulation of free Gly_{10} using ABSINTH model[60] implemented in CAMPARI package to calculate the conformational properties. ABSINTH has been highly successful in describing the IDP sequence composition-structural preference relationship[60, 63, 200, 285], and thus provides a good reference for examining the conformational properties of Gly_{10}. The MC simulation follows the protocol as defined in "file abs3.2_opls.prm" in CAMPARI. The peptide was put in a droplet of 80 Å radius and simulated at 300 K

under NVT conditions. The simulation lasted 41,000,000 MC steps, with the first 1,000,000 steps excluded in the analysis. A 200-ns simulation of Gly_{10} was also performed in the GBSW implicit solvent for comparison.

• *MSES simulations:* In MSES simulations of $(AAQAA)_3$, AT and CG models were coupled using U_{MSES} imposed on (i, i+4) Cα atoms that form native helical contacts. U_{MSES} with soft asymptote was used to reduce energy penalty for significantly different conformations between AT and CG models and avoid exchange bottleneck between neighboring replicas[243],

$$U_{MSES}(r_{AT}, r_{CG}) = \sum_i 0.5k_i\left(d_i^{AT} - d_i^{CG}\right)^2, \qquad if\ |\Delta d_i| \le d_s$$

$$= \sum_i A + \frac{B}{\left(d_i^{AT} - d_i^{CG}\right)^s} + f_{max}\left(d_i^{AT} - d_i^{CG}\right), \qquad if\ |\Delta d_i| > d_s$$

$$with\ \Delta d_i = d_i^{AT} - d_i^{CG} \tag{5.11}$$

where $d_{\mathrm{i}}^{\mathrm{AT}}$ and $d_{\mathrm{i}}^{\mathrm{CG}}$ are Cα-Cα distances for residue pair *i* in atomistic and CG models, respectively. d_s is the distance threshold where U_{MSES} smoothly switches from quadratic to the soft asymptote. *s* determines how quickly the maximal force, f_{max}, is approached at large distance difference $\left|d_{\mathrm{i}}^{\mathrm{AT}} - d_{\mathrm{i}}^{\mathrm{CG}}\right|$. A and B are identified by requiring both MSES coupling energy and force to be continuous at $\left|d_{\mathrm{i}}^{\mathrm{AT}} - d_{\mathrm{i}}^{\mathrm{CG}}\right| = d_{\mathrm{s}}$. The H-REX protocol previously optimized for disordered proteins[243] is used in this work. For comparison, three independent sets of simulations were preformed using T-REX in GBSW implicit solvent and MSES with the GBSW model coupled to the Cα-only Gō-like model (MSES/Gō) or the HyRes model (MSES/HyRes). For each protocol, both control and folding

simulations were initiated from helical and extended structures, respectively, for convergence analysis. All simulations were performed with 16 replicas and temperatures ranging from 270 K to 500 K. Other simulation details are listed in Table 5.1.

Table 5.1 T-REX and MSES Simulations of $(AAQAA)_3$.

Protocol	Run	Simulation time (ns/replica)	Exchange acceptance ratio
T-REX	Control	100	0.61
	Folding	100	0.61
MSES/Gō	Control	50	0.55
	Folding	50	0.55
MSES/HyRes	Control	50	0.51
	Folding	50	0.51

5.3 Results and Discussions

5.3.1 Side chain bonded parameter optimization

Packing of peptide side chains has long been recognized as a key factor in determining protein structure (and folding)[286]. A flexible CG protein model thus needs to provide a reasonably realistic description of side chain geometries. Accordingly, the equilibrium bond and angle parameters (as well as vdW radii) of the HyRes model are assigned to maximally reproduce atomistic distributions according the mapping scheme (Table 5.6). The corresponding bond and angle force constants are assigned to reproduce the root-mean-squared (RMS) fluctuations from atomistic simulations. As summarized in Figure 5.2, the average values of all virtual bond lengths and angles observed in the atomistic GBSW model are well reproduced in HyRes. Importantly, the CG model is also able to reproduce the distributions of most virtual bonds and angles, except those displaying multi-modal distributions (Figure 5.12 and Figure 5.13). These are the cases where the side chain samples

multiple rotamers at the atomic resolution and reflect an inherent consequence of coarse-graining. For these bonds and angles, the associated standard deviation derived from atomistic simulations are larger than those in the HyRes model (Figure 5.2). Once the bond and angle parameters are assigned, side chain χ (N_i-CA_i-CB_i-CC_i) dihedral potentials are parameterized by fitting atomistic distributions to cosine expansions for residues Lys, Arg, His, Phe, Tyr and Trp. The resulting parameters are summarized in Table 5.8. As shown in Figure 5.14, all atomistic distributions along χ can be well reproduced by HyRes. Additional dihedral potentials are imposed on ψ' (CB_i-CA_i-C_i-O_i), to finely tune the helical propensities of individual amino acids (independent of the backbone φ/ψ CMAP cross-term). Since Asp, Asn, Thr, Cys and Val have been suggested to have lower helical propensities than other non-Gly, non-Pro residues[287, 288], weak U_{dihedral} along ψ' is applied to these residues to capture this feature (see Table 5.9).

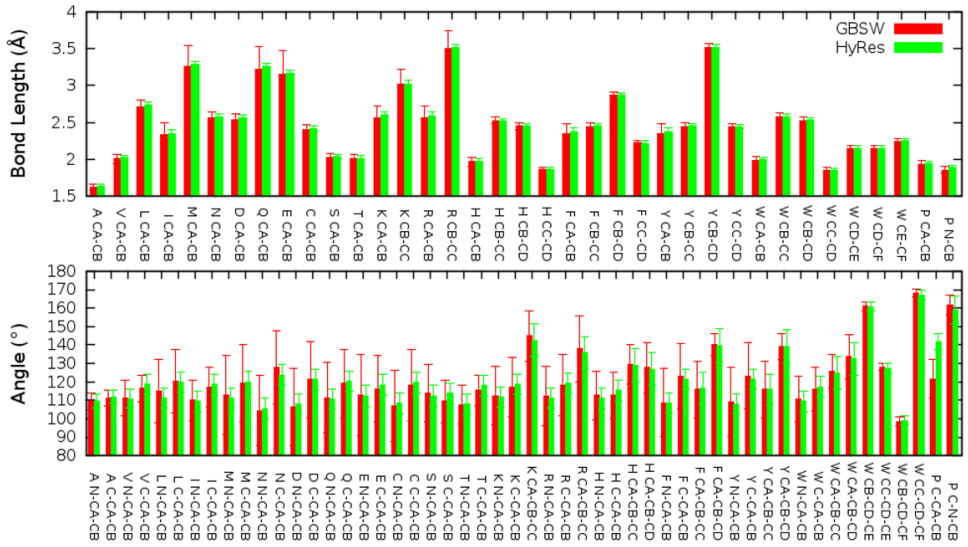


Figure 5.2 Averaged values and standard deviations of bond lengths (top) and angles (bottom) for all 20 amino acids, derived from 300 ns of GBSW and 100 ns HyRes simulations of dipeptides at 300 K.

5.3.2 Scaling of backbone vdW interactions

The original CHARMM19 vdW parameters for the backbone over-estimate the effective strengths of nonspecific dispersion interactions in HyRes due to a lack of solvent screening, and this leads to severe over compaction in peptide simulations. To address this limitation, we uniformly scale down the strengths of backbone vdW interactions (ε_i in Eq. 5.9). Determination of the appropriate scaling factor is guided by explicit solvent simulations of Gly_{10}, a highly flexible peptide that does not have any side chains and is thus particularly sensitive to the balance of water-mediated backbone-backbone interactions. Two sets of explicit solvent simulations of Gly_{10} were performed on representative "compact" and "extended" states to estimate the ratio between net contribution of vdW interactions and the intra-peptide component alone:

$$f = \frac{U_{vdw}^{tot}(cmp) - U_{vdw}^{tot}(ext)}{U_{vdw}^{intra}(cmp) - U_{vdw}^{intra}(ext)} \tag{5.12}$$

The results show that the net contribution of vdW interactions to the stabilization of compact states is about 20-40% of the intra-peptide component alone (see Table 5.10 for details). That is, solvent screening reduces the energetic contribution of nonspecific vdW interactions to conformational collapse by 60-80%. Accordingly, vdW interactions of all backbone atoms (ε_i) are reduced to 30% of the original CHARMM19 values to approximately account for solvent screening in HyRes. The scaling significantly reduces over-compaction. The conformational properties of Gly_{10} calculated using HyRes agrees well with results from various atomistic simulations in implicit and explicit solvents (Table 5.2 and Figure 5.3). We note that GBSW is known to generate overly compact ensembles (e.g., compared to

ABSINTH, blue vs green traces in Figure 5.3). HyRes appears to display a small bias towards compaction, which is likely due to lack of more sophisticated implicit treatment of solvent. Benchmark simulations suggest that the compaction bias is smaller for more complex and longer peptides (see below), and the current choice of scaling appears appropriate.

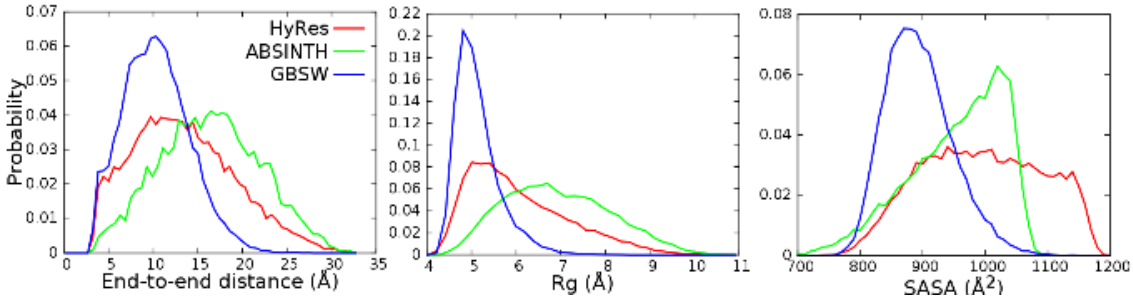


Figure 5.3 Probability distribution of the end-to-end distance, R_g and SASA of Gly_{10} obtained from HyRes, GBSW and ABSINTH simulations at 300 K.

Table 5.2 Ensemble averaged end-to-end distance (d_{e2e}), radius of gyration (R_g), and solvent accessible surface area (SASA) of Gly_{10} simulated using HyRes, GBSW, CHARMM27, CHARMM36, Amber ff12SB and ABSINTH.

Model	$\langle d_{e2e} \rangle$ (Å)	$\langle R_g \rangle$ (Å)	$\langle$SASA$\rangle$ (Å^2)
HyRes	13.65	6.15	1000
GBSW	10.65	5.23	899
CHARMM27[§]	13.30	5.65	880
CHARMM36[§]	18.14	7.22	1010
Amber ff12SB[§]	14.78	6.31	934
ABSINTH	16.54	6.94	959

[§] Results are from Drake and Pettitt's work[289].

5.3.3 Side chain non-bonded parameters

The vdW radii of all amino acid side chain beads were first assigned based on the average volume occupied by corresponding atoms in GBSW implicit solvent simulations of dipeptides, as $R = \sqrt[3]{3V/4\pi}$. The initial values of vdW radii were fine-tuned by examining the backbone ϕ/ψ adiabatic energy surfaces of dipeptides and PMFs of side chain pair-wise interactions. The backbone ϕ/ψ energy surface is

particularly sensitive to the covalent geometry and vdW radii. Other non-bonded parameters such as hydrogen bonding interactions and vdW interaction strengths have relatively weaker effects. It was found that special 1-4 vdW radii were needed for most amino acid side chain beads to reproduce results from GBSW implicit solvent, e.g., comparing Figure 5.15 and Figure 5.16. The final assignment of vdW radii is summarized in Table 5.3. MD simulations of Ala dipeptide were also performed to compare HyRes with AMBER ff99SB-ILDN[74] and CHARMM36m[73] explicit solvent force fields. The results are shown in Figure 5.4. Although details in free energy landscapes are different among all three models, probabilities of visiting various secondary structure regions (e.g., α helical vs extended states) appear similar. Note that inherent secondary structure propensities of HyRes have also been fined tuned using the ϕ/ψ CMAP and ψ' torsion terms based on model peptide simulations. This will be described in detail in the following sections.

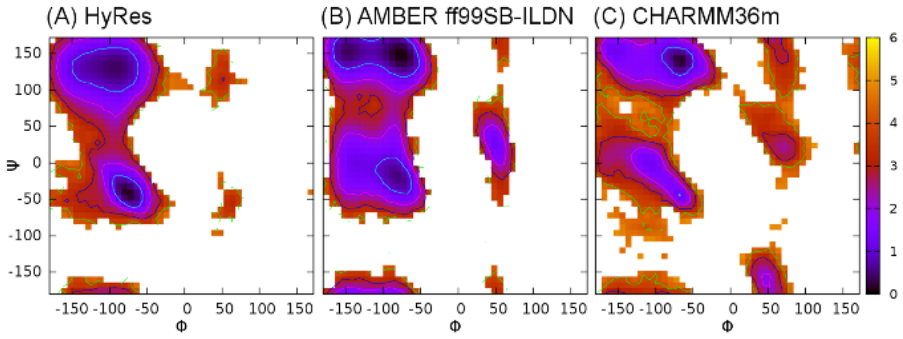


Figure 5.4 Backbone ϕ/ψ free energy profiles of alanine dipeptide obtained from 200 ns of MD simulations using (A) HyRes, (B) AMBER ff99SB-ILDN, and (C) CHARMM36m.

To determine vdW interaction strengths of side chain beads, we followed a deconvolution strategy previously described by Bereau and Deserno[257] to reproduce the 20 × 20 MJ matrix using 20 parameters (ϵ_i') based on the Lorentz-Berthelot (LB) mixing rule (Eq. 5.9). The final values of 20 ϵ_i' were listed in Table 5.4. As shown in Figure 5.5, the pair-wise vdW interaction strengths calculated using 20 ϵ_i' is in excellent agreement with the original MJ matrix, with the overall RMSD of 0.25 RT. Only 11 out of the 210 pair-wise interactions deviate by greater than 0.5 RT (Table 5.5). Deviation of Cys-Cys pair-wise interaction from MJ potential may be related to disulfide bond formation in proteins. Among these 11 pairs of residues, nine pairs are between charged residues or between charged and hydrophobic residues. Interactions between oppositely charged residues were underestimated by this deconvolution method, while overestimated for residues with same charges (Table 5.5). When explicit electrostatic interactions are taken into account between charged residues, the actual strengths of interactions involving charged residues agree better with the MJ matrix (Figure 5.17). Importantly, the nonspecific attractive interactions between oppositely charged residues in HyRes are ~2 kcal/mol (Figure 5.17), which is very similar to results obtained from atomistic simulations in explicit solvent[266, 290].

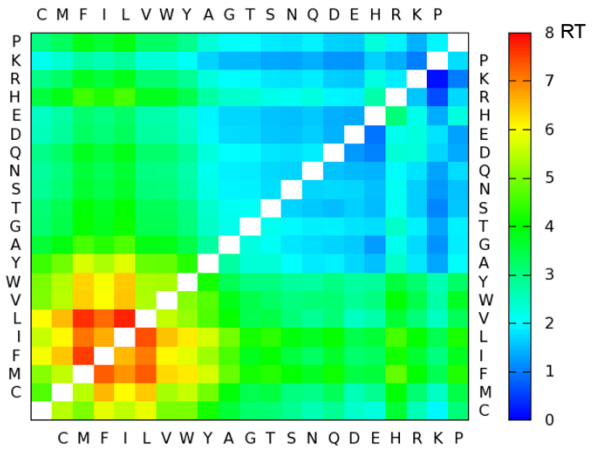


Figure 5.5 $\epsilon_{i,j}$' parameters from original MJ statistical potentials (upper left) and in HyRes model by calculating $\epsilon_{i,j}$' from ϵ_{i},' using the LB mixing rule (lower right).

The deconvoluted parameters (ϵ_i') for 20 amino acids are then rescaled based on the PMFs of pair-wise interactions between side chain analogs using atomistic results as a reference. During the rescaling process, the minimum and maximum values of ϵ_i' (for Lys and Leu, respectively) are adjusted while the relative scale $(\epsilon_i' - \min[\epsilon_i'])/(\max[\epsilon_i'] - \min[\epsilon_i'])$ of each residue is maintained. Optimal agreement with atomistic PMFs were found with $\min[\epsilon_i']$ = 0.1 kcal/mol and $\max[\epsilon_i']$ = 0.9 kcal/mol. For residues with multiple CG beads in the side chain, such as Lys, Arg, His, Phe and Trp (see Figure 5.1), the scaled ϵ_i' needs to be further partitioned to the constituent CG beads based on the mapping scheme (Table 5.6). The final parameters are summarized in Table 5.3. As illustrated in Figure 5.6**,** the resulting PMFs of side chain interactions in HyRes are in excellent agreement with atomistic results. Note that HyRes tends to shift the minima of PMFs towards larger values, which is a direct consequence of using CG beads to represent groups of atoms with anisotropic structures. An implication is that HyRes likely remains limited in accurate description of side chain packing. However, the main objective

of developing HyRes is to provide a better CG modeling for MSES simulation of atomistic ensembles. A highly precise of long-range interactions at the CG level is not expected to be critical.

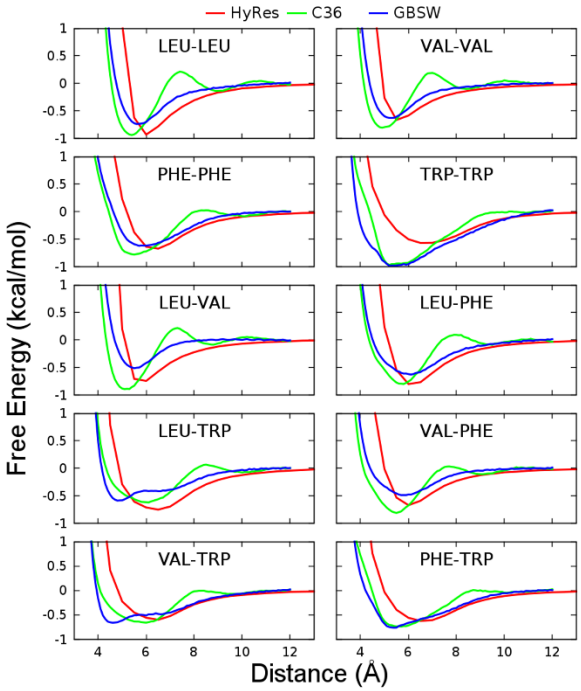


Figure 5.6 Free energy profiles as a function of separation distance between the center of mass of hydrophobic amino acid side chain analogs.

Table 5.3 Lennard-Jones parameters for all side chain beads.

Residue	CG bead	ε_i (kcal/mol)	$r_i^{min}/2$ (Å)	$r_i^{min}/2$ (1-4) (Å)
Ala	CB	-0.308	2.12	2.12
Val	CB	-0.62	2.75	2.75
Leu	CB	-0.9	2.96	2.96
Ile	CB	-0.772	2.97	2.97
Met	CB	-0.636	2.98	3.68
Asn	CB	-0.18	2.65	3.25
Asp	CB	-0.148	2.61	3.11
Gln	CB	-0.204	2.89	3.89
Glu	CB	-0.14	2.85	3.95
Cys	CB	-0.532	2.47	2.77
Ser	CB	-0.188	2.32	2.72
Thr	CB	-0.228	2.62	2.62
Pro	CB	-0.212	2.77	2.77
Lys	CB	-0.05	2.78	3.48
Lys	CC	-0.05	2.36	3.06
Arg	CB	-0.135	2.78	3.18
Arg	CC	-0.135	2.54	2.94
His	CB	-0.108	2.34	2.64
His	CC	-0.081	2.18	2.48
His	CD	-0.081	2.11	2.41
Phe	CB	-0.22	2.64	2.94
Phe	CC	-0.22	2.33	2.63
Phe	CD	-0.22	2.33	2.63
Tyr	CB	-0.197	2.64	2.94
Tyr	CC	-0.197	2.33	2.63
Tyr	CD	-0.0984	2.45	2.75
Trp	CB	-0.168	2.42	2.72
Trp	CC	-0.084	2.24	2.54
Trp	CD	-0.168	2.09	2.39
Trp	CE	-0.168	2.33	2.63
Trp	CF	-0.168	2.33	2.63

Table 5.4 ϵ_i' (in RT) before rescaling obtained from deconvoluting the MJ matrix.

AA(i)	C	M	F	I	L	V	W
ϵ_i'	4.63	5.51	7.53	6.65	7.73	5.37	5.31
AA(i)	Y	A	G	T	S	N	Q
ϵ_i'	4.3	2.75	2.14	2.08	1.74	1.67	1.87
AA(i)	D	E	H	R	K	P	
ϵ_i'	1.4	1.34	2.68	1.87	1	1.94	

Table 5.5 Residue pairs with the deviation $\Delta\epsilon_{i,j}' = \epsilon_{i,j}'(MJ) - \epsilon_{i,j}'(LB)$ greater than 0.5 RT.

AA_i	C	F	L	W	Y	D
AA_j	C	K	K	P	K	R
$\Delta\epsilon'_{i,j}$	0.81	0.62	0.59	0.52	0.53	0.67
AA_i	D	E	E	R	K	
AA_j	K	R	K	K	K	
$\Delta\epsilon'_{i,j}$	0.5	0.69	0.64	-0.78	-0.88	

5.3.4 Tuning of electrostatic interactions

The Debye–Hückel-type potential has proven highly successful in description nonspecific protein interactions[291]. A main parameter subject to tuning in Eq. 5.10 is the effective dielectric constant ε_r, which directly scales the strengths of electrostatic interactions. The optimal value for a given CG model should be chosen to provide a proper balance among various competing short- and long-range interactions. We evaluate such balance by examining the conformational properties of a set of 30 variants of $(EK)_{25}$ peptide, which were originally designed by Das and Pappu[64] to study how charge patterning governs the conformational preferences of IDPs. A parameter, κ, ranging from 0 to 1, has been defined to quantify the charge patterning, with low values for well-mixed sequences and high values for segregation of oppositely charged residues in the sequence. As shown Figure 5.7, the ensemble averaged radius of gyration (R_g) of $(EK)_{25}$ peptide decreases as κ increases, a trend that is well reproduced by HyRes with $\varepsilon_r = 20$. This suggests that the overall balance between electrostatic and other interactions is reasonable. Nonetheless, HyRes tends to generate conformational ensembles more compact than ABSINTH for some sequences with intermediate κ (e.g., between 0.1 and 0.4). This

again likely reflects inherent limitations associated with CG representations and greatly simplified energy function employed (particularly, treatment of solvation).

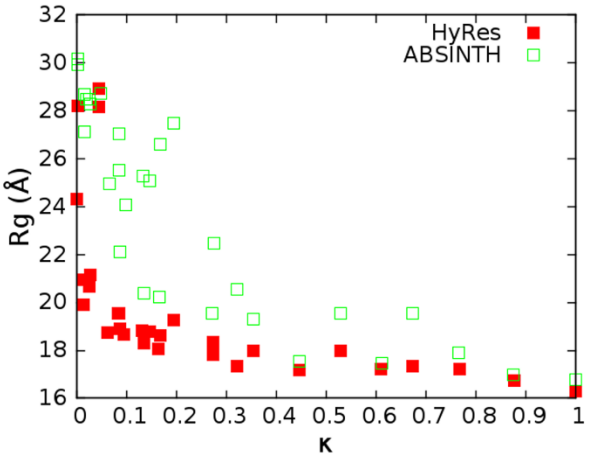


Figure 5.7 Averaged R_g as a function of charge patterning parameter κ for 30 variants of $(EK)_{25}$ peptide. Values from HyRes simulations are uniformly shifted up by 4 Å to roughly account for larger side chain beads. ABSINTH results are taken from Ref. [64].

5.3.5 Balancing backbone hydrogen-bonding and torsion potentials

Once all vdW parameters were assigned, the strength of backbone hydrogen bonding interactions was tuned together with the ϕ/ψ CMAP term to recapitulate the secondary structure propensities of model peptides. In this work, we mainly focused on helix-coil transitions and used $(AAQAA)_3$ as a primary model peptide in backbone tuning. $(AAQAA)_3$ is partially helical that is about 20% folded at 300 K[292]; it has been frequently used in atomistic and CG force field optimization[59, 243, 293]. The backbone parameters optimized based on $(AAQAA)_3$ were then fine-tuned using KID, a slightly more complex IDP that folds into two helices upon binding to specific target KIX[283]. NMR secondary chemical shifts and atomistic simulations have shown that N-terminal helical segment is about 50-60% folded, while the second helical segment is only 10-15% folded[61, 294]. The final CMAP

is shown in Figure 5.18, and hydrogen bond interaction strength (ε_{HB} in Eq. 5.7) is set to 1.5 kcal/mol. Figure 5.8 depicts the residual helicity profiles of $(AAQAA)_3$ and KID calculated from the HyRes model, with representative structures shown in Figure 5.19 and Figure 5.20, respectively. The results are highly consistent with both NMR and previous atomistic simulations, suggesting that the HyRes model is able to capture the sequence dependence of helical propensity quite well. This property is expected to be highly beneficial for driving faster atomistic sampling in MSES simulations of IDPs.

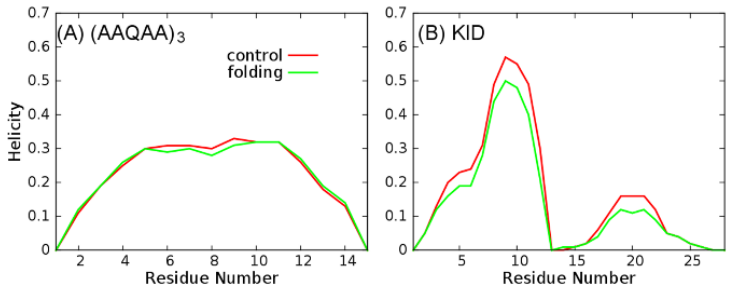


Figure 5.8 Residual helicity profiles of (A) $(AAQAA)_3$ and (B) KID at 300 K calculated using the HyRes model.

5.3.6 HyRes simulations of nontrivial IDPs

The HyRes model was used to study the RS peptide, a highly charged, disordered peptide that has been extensively characterized by NMR[295], small angle X-ray scattering (SAXS)[296], and various atomistic simulations[72, 73, 296]. RS peptide in HyRes simulations showed averaged R_g of 10.1 Å at 300 K (Figure 5.21), which agreed well with SAXS measurements (12.62 Å)[296] by considering CG bead radii of 2-3 Å (Table 5.3). The optimized HyRes model was further validated using two longer IDPs with more complex conformational properties, ACTR and NCBD. These two IDPs can synergistically fold into a stable helical complex upon binding[218]. In the unbound state, ACTR has a very low level of residual helices,

while NCBD contains has molten globule characteristics with folded-like helical contents[297, 298]. Key conformational properties obtained from T-REX simulations in HyRes are summarized in Figure 5.9. The results show that HyRes yields residual helices in the correct regions revealed in previous studies[81]. It also successfully predicts higher helical contents in NCBD than ACTR, even though the model over-estimates the helical content of ACTR while under-estimating that of NCBD. The calculated R_g distribution of ACTR also agrees well with the result from ABSINTH simulations. However, the overall dimension of ACTR seems to be under-estimated compared to SAXS measurements, which yielded an average R_g ~26.3 Å at 278 K for a longer 71-residue construct[299] (vs. the 47-residue segment simulated here). Over-estimation of the helical content is likely a key contributing factor here. For the larger peptide NCBD (59 residues), HyRes generates largely globular conformational ensembles with an averaged R_g of ~11.4 Å, which is in reasonable agreement with the value of ~15.2 Å derived from SAXS (after taking into account CG bead radii of 2-3 Å in HyRes, see Table 5.3)[26]. For a β-hairpin peptide GB1m3, which is ~86% folded at 298 K based on NMR measurement[282], the current HyRes model could sample the folded state, but underestimates its stability (Figure 5.22). Taken together, these benchmark simulations support the notion that HyRes has achieved the main design objective, which is to provide a semi-quantitative description of secondary structures and a qualitative description of the long-range conformational properties. We note that the model could be further improved by including better treatment of hydrogen bonds and solvation.

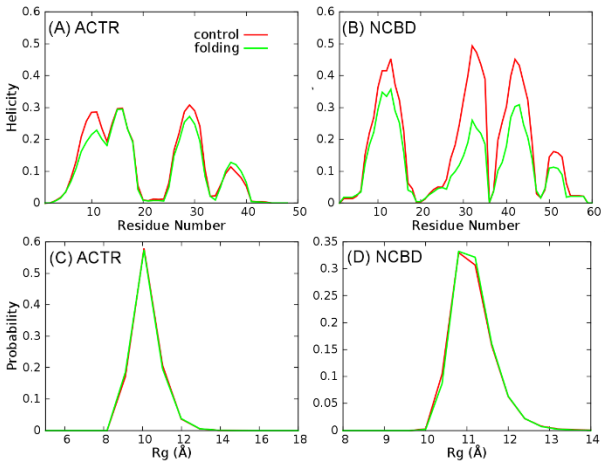


Figure 5.9 Averaged residue helicity profiles and distributions of Rg of ACTR and NCBD calculated using HyRes.

5.3.7 Efficiency of HyRes in driving helix-coil transitions in MSES simulations

Previous work has shown that MSES approach is able to accelerate sampling of protein conformations in the rough, complex atomistic energy landscapes, even when the atomistic model is only coupled with a simple Cα-based Gō-like model[83, 243]. MSES simulations of disorder proteins should benefit further from CG models that can generate both local and long-range structural fluctuations consistent with those observed in atomistic models. The performance of the newly developed HyRes model in MSES simulations is examined using the model peptide $(AAQAA)_3$. As shown in Figure 5.23, atomistic helicity profiles obtained from T-REX and MSES simulations, coupled to either the Cα-only Gō-like model or HyRes, are essentially identical as expected. This validates that MSES simulations are able to generate proper canonical ensemble free of bias due to MSES coupling. The efficacy of using HyRes to accelerate atomistic conformational transitions in MSES simulations is then assessed by computing the reversible helix/coil transition rate for each residue in $(AAQAA)_3$ at the atomistic level. As summarized in Figure 5.10, the atomistic

helix/coil transition rate observed in MSES/Gō simulation is ~2 fold of that in atomistic T-REX simulations. Using the HyRes model in MSES simulations further accelerates atomistic structural transitions to ~3 fold (red traces in Figure 5.10). Although the exact values of helix-coil transition rate depend on a lot of factors, such as capping groups, peptide length[300], concentration, temperature[301], hydrodynamics[302, 303], etc., our result demonstrates that the HyRes model could drive faster conformational transitions in atomistic energy landscapes. Faster helix-coil transitions translate well into faster convergence rate in averaged conformational properties such as residue helicity profiles. As shown in Figure 5.11, the residual helicity profiles of $(AAQAA)_3$ obtained from MSES/HyRes simulations converge rapidly in both control and folding runs. Especially in the folding runs, MSES/HyRes simulation became converged faster than both MSES/Gō and T-REX simulations.

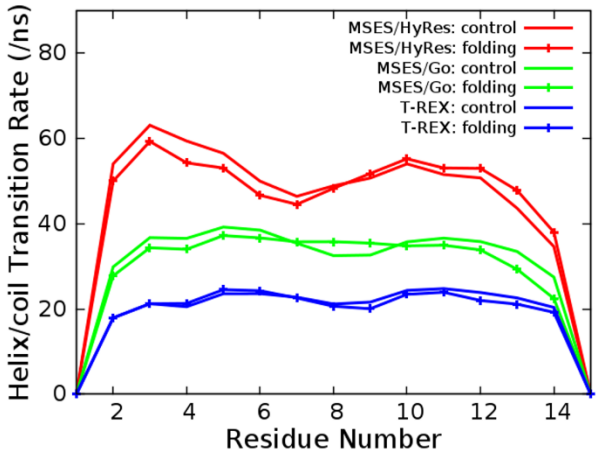


Figure 5.10 Reversible helix-coil transition rates for each residue of $(AAQAA)_3$ at atomistic level. For each replica, we first computed helix/coil transition rate as number of reversible helix/coil transitions divided by simulation time. Results for each simulation shown here are averaged values over all replicas. Note that only the second half of each trajectory was used in these calculations.

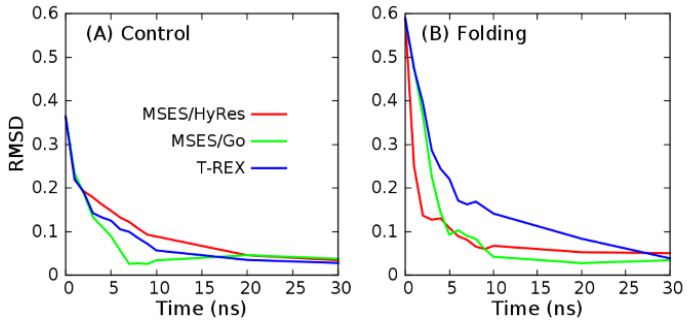


Figure 5.11 Convergence rate of $(AAQAA)_3$ residual helicity profiles obtained from MSES/HyRes, MSES/Gō, and T-REX simulations. At each time point *t*, we first computed residual helicity profile using trajectory from the beginning to time *t*. RMSD(t) is calculated between the profile at time *t* and the averaged one calculated from all six simulations shown in Figure 5.23.

5.4 Conclusions

Understanding the structural basis of IDPs recognition and regulation is critical in unveiling the mechanisms of IDP functions and associated diseases. Previous studies have demonstrated that the MSES scheme is highly suitable for characterizing conformational properties of IDPs. One of the factors that limits the MSES sampling efficiency is the ability of CG models to generate both local and long-range structural fluctuations that are largely consistent with AT models. Cα-only Gō-like models have proven to be inadequate. Here, we describe a new HyRes model designed especially for MSES simulations of IDPs. The model treats solvent molecules implicitly. The peptide backbone is at atomistic resolution, which allows for semi-quantitative description of secondary structure propensities of IDPs. Side chains are represented at an intermediate resolution with one or more CG beads, whose parameters were carefully tuned provide a qualitative description of transient peptide-peptide interactions and long-range conformational properties. Benchmark simulations of several IDPs with various level of complexities demonstrate that the

HyRes model has achieved all design objectives. Initial application to simple model IDPs suggests HyRes can drive faster atomistic transitions in MSES and leads to faster convergence in atomistic structure ensembles. With further optimization of this CG model using more peptides containing β strands, the HyRes model could be extended to study IDPs with more complex secondary structures. We also note that achieving the best sampling efficiency with coupling to HyRes will require further optimization of the MSES coupling scheme and H/T-REX protocol. These developments are ongoing, which can be anticipated to further improve the efficiency of MSES for studies of larger and more complex IDPs frequently involved in cellular signaling and regulation.

5.5 Supplementary Information

Table 5.6 Mapping of side chain atoms from the CHARMM 22 atomistic model in HyRes. A single CB bead is used in all the other residues (except Gly).

Residue	CG bead	Atom names in CHARM 22
Lys	CB	CB HB1 HB2 CG HG1 HG2 CD HD1 HD2
	CC	CE HE1 HE2 NZ HZ1 HZ2 HZ3
Arg	CB	CB HB1 HB2 CG HG1 HG2 CD HD1 HD2
	CC	NE HE CZ NH1 HH11 HH12 NH2 HH21 HH22
His	CB	CB HB1 HB2 CG
	CC	CD2 HD2 NE2
	CD	ND1 HD1 CE1 HE1
Phe	CB	CB HB1 HB2 CG CD1 HD1
	CC	CD2 HD2 CE2 HE2
	CD	CE1 HE1 CZ HZ
Tyr	CB	CB HB1 HB2 CG CD1 HD1
	CC	CD2 HD2 CE2 HE2
	CD	CE1 HE1 CZ OH HH
Trp	CB	CB HB1 HB2 CG
	CC	CD1 HD1 NE1 HE1
	CD	CD2 CE2
	CE	CZ2 HZ2 CH2 HH2
	CF	CE3 HE3 CZ3 HZ3

Table 5.7 Sequences of model peptides used in this work. All peptides were capped with an acetyl group at N-terminus and N-methyl amide at C-terminus.

Peptide	**κ**	**Sequence**
Gly_{10}		GGGGG GGGGG
$(AAQAA)_3$		AAQAA AAQAA AAQAA
KID		TDSQK RREIL SRRPS YRKIL NDLSS DAP
ACTR		EGQSD ERALL DQLHT LLSNT DATGL EEIDR ALGIP ELVNQ GQALE PK
NCBD		PNRSI SPSAL QDLLR TLKSP SSPQQ QQQVL NILKS NPQLM AAFIK QRTAK YVANQ PGMQ
RS		GAMGP SYGRS RSRSR SRSRS RSRS
GB1m3		KKWTY NPATG KFTVQ E
$(EK)_{25}$	0.0009	EKEKE KEKEK EKEKE KEKEK EKEKE KEKEK EKEKE KEKEK EKEKE KEKEK
	0.0025	EEEKK KEEEK KKEEE KKKEE EKKKE EEKKK EEEKK KEEEK KKEEE KKKEK
	0.0139	KEKKK EKKEE KKEEK EKEKE KEEKK KEEKE KEKEK KKEEK EKEEK KEEEE

	0.0140	KEKEK KEEKE KKEEE KKEKE KEKKK EEKKK EEKEE KKEEK KKEEK EEEKE
	0.0245	KEKEE KEKKK EEEEK EKKKK EEKEK EKEKE EKKEE KKKKE EKEEK EKEKE
	0.0273	EEEKK EKKEE KEEKK EKKEK EEEKK KEKEE KKEEE KKKEK EEEEK KKKEK
	0.0450	EEEEK KKKEE EEKKK KEEEE KKKKE EEEKK KKEEE EKKKK EEEEK KKKEK
	0.0450	KKKKE EEEKK KKEEE EKKKK EEEEK KKKEE EEKKK KEEEE KKKKE EEEKE
	0.0624	EEKKE EEKEK EKEEE EEKKE KKEKK EKKKE EKEKE KKKEK KKKEK EEEKE
	0.0834	EKKKK KKEEK KKEEE EEKKK EEEKK KEKKE EKEKE EKEKK EKKEE KEEEE
	0.0841	EKEKK KKKEE EKKEK EEEEK EEEEK KKKKE KEEEK EEKKE EKEKK KEEKK
	0.0864	EKKEE EEEEK EKKEE EEKEK EKKEK EEKEK KEKKK EKKEE EKEKK KKEKK
	0.0951	KEKKK EKEKK EKKKE EEKKK EEEKE KKKEE KKEKK EKKEE EEEEE KEEKE
	0.1311	EKKEK EEKEE EEKKK KKEEK EKKEK KKKEK KKKKE EEEEE KEEKE KEKEE
	0.1354	KKEKK EKKKE KKEKK EEEKE KEKKE KKKKE KEKKE EEEEE EEKEE KKEEE
	0.1458	EKEKE EKKKE EKKKK EKKEK EEKKE KEKEK KEEEE EEEEE KEKKE KKKKE
	0.1643	EKEKK KKKKE KEKKK KEKEK KEKKE KEEEK EEKEK EKKEE KKEEE EEEEE
	0.1677	KEEKK EEEEE EEKEE KKKKK EKKKE KKEEE KKKEE KKKEE EEEEK KKKEK
	0.1941	EEEEE KKKKK EEEEE KKKKK EEEEE KKKKK EEEEE KKKKK EEEEE KKKKK
	0.2721	EEKEE EEEEK EEEKE EKKEE EKEKK EKKEK EEKKE KKKKK KKKKK KKEEE
	0.2737	EEEEE EEEEK EKKKK KEKEE KKKKK KEKKE KKKKE KKEEE EEEKE EEKKK
	0.3218	KEEEE KEEKE EKKKK EKEEK EKKKK KKKKK KKKEK KEEEE EEEEK EKEEE
	0.3545	EEEEE KEEEE EEEEE EEKEE KEKKK KKKEK KKKKK KEKEK KKKEK KEEKK
	0.4456	EEEEK EEEEE KEEEE EEEEE EEEKK KEEKK KKKEK KKKKK KEKKK KKKKK
	0.5283	EEEEE EEEEE EKEEE EKEEK EEKEK KKKKK KKKKK KKKKK KKEEK KEEKE
	0.6101	KEEEE EEEKE EKEEE EEEEE EKEEE EKEEK KKKKK KKKKK KKKKK KKKKE

	0.6729	KKEKK KEKKE EEEEE EEEEE EEEEE EEEEK EEKKK KKKKK KKKKK KKEKK
	0.7666	EKKKK KKKKK KKKKK KKKKK KKEEE EEEEE EEEEE EEEEE KKEEE EEKEK
	0.8764	KEEEE KEEEE EEEEE EEEEE EEEEE EEKKK KKKKK KKKKK KKKKK KKKKK
	1.0000	EEEEE EEEEE EEEEE EEEEE EEEEE KKKKK KKKKK KKKKK KKKKK KKKKK

Table 5.8 Parameters of $U_{dihedral}$ for side chain χ (N_i-CA_i-CB_i-CC_i).

Residues	k_χ(kcal/mol)	n	δ
Lys	0.3	3	0
Lys	0.5	1	75
Arg	0.3	3	0
Arg	0.5	1	75
His	0.6	3	0
His	0.4	1	75
His	0.1	1	50
Phe	0.8	3	180
Phe	0.3	1	100
Phe	0.1	1	50
Tyr	0.9	3	180
Tyr	0.3	1	90
Tyr	0.1	1	50
Trp	0.8	3	0
Trp	0.8	1	85

Table 5.9 Parameters of $U_{dihedral}$ applied to dihedral ψ' (CB_i-CA_i-C_i-O_i)

Residues	$k_{\psi'}$(kcal/mol)	n	δ
Asp	0.3	1	240
Asn	0.3	1	240
Thr	0.3	1	240
Cys	0.3	1	240
Val	0.3	1	240

Table 5.10 vdW interaction energies (in kcal/mol) for Gly_{10} in representative compact and extended states derived from two independent sets of 10-ns explicit solvent simulations. The compact and extended states are mimicked by restraining the peptide end-to-end distances to 7 and 20 Å, respectively.

Simulation	Energy term	U_{cmp}	U_{ext}	U_{cmp} - U_{ext}
1	$U_{vdw}^{intra-pept}$	-1.78	0.33	-2.11
	U_{vdw}^{inter}	-46.69	-49.84	3.15
	$U_{vdw}^{intra-solv}$	8358.30	8359.74	-1.44
	U_{vdw}^{tot}	8309.82	8310.22	-0.40
2	$U_{vdw}^{intra-pep}$	-0.78	-0.42	-0.36
	U_{vdw}^{inter}	-47.07	-48.91	1.85
	$U_{vdw}^{intra-solv}$	8357.60	8359.23	-1.62
	U_{vdw}^{tot}	8309.76	8309.90	-0.14

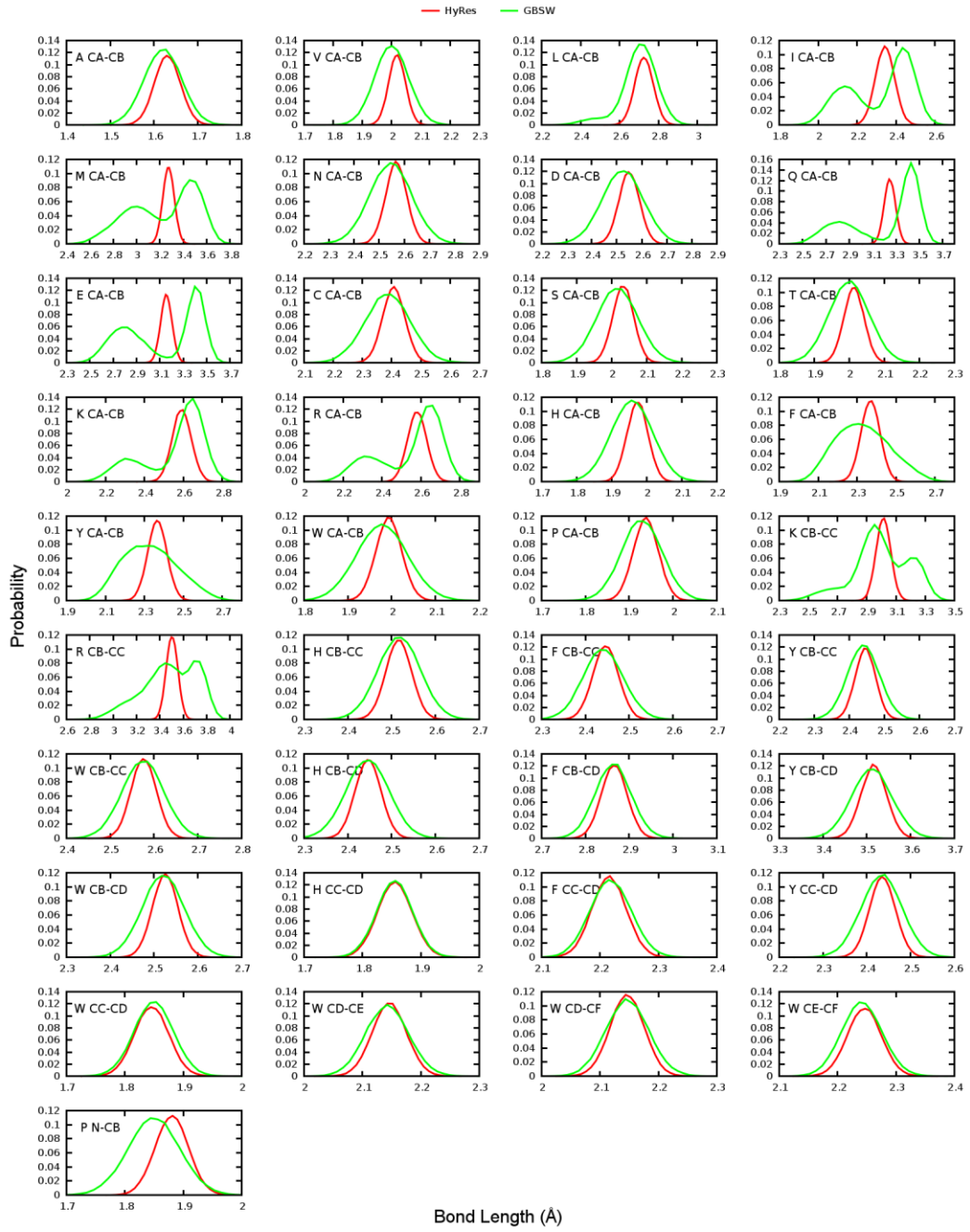


Figure 5.12 Distributions of side chain virtual bond lengths from MD simulations of dipeptides using the GBSW and HyRes models.

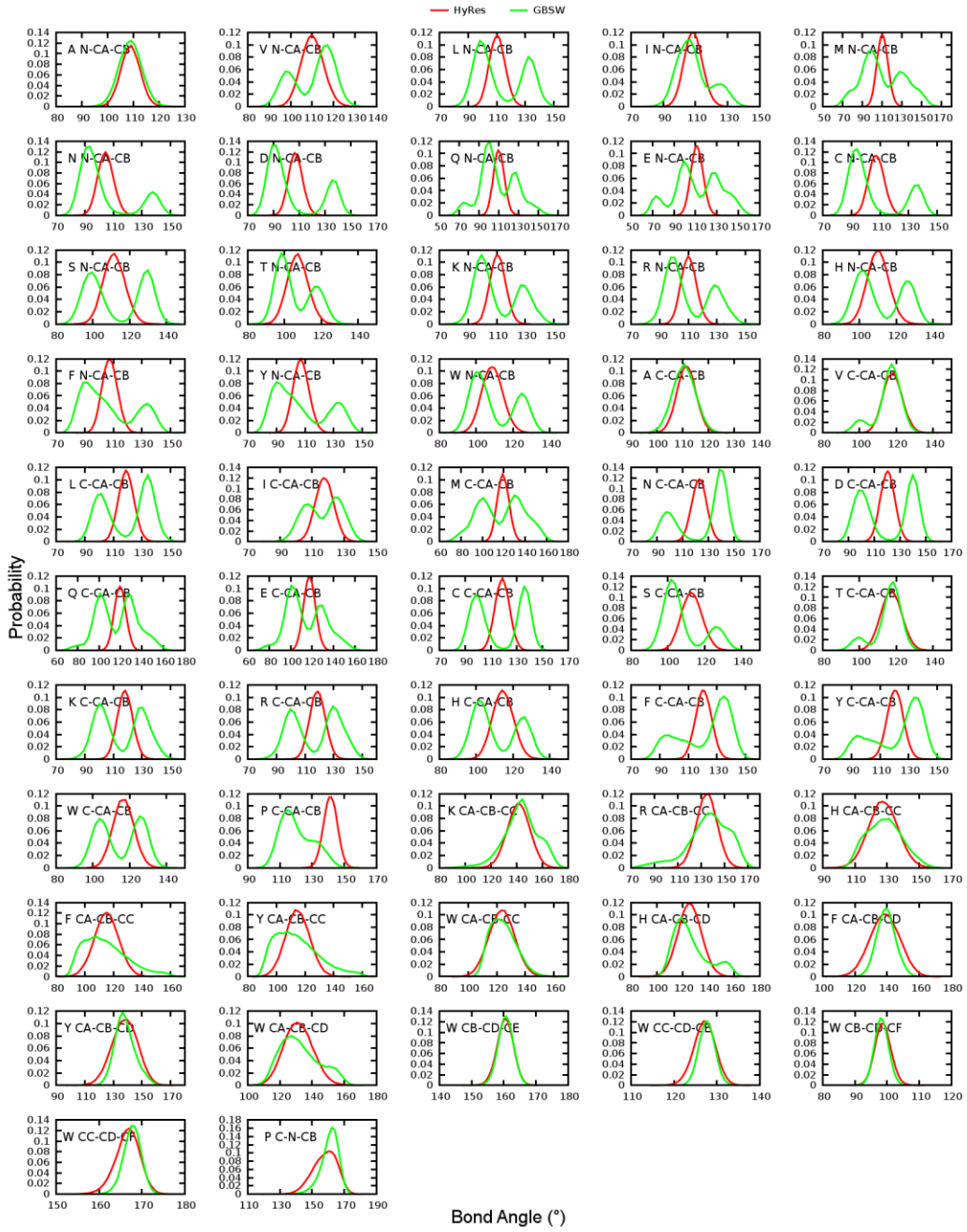


Figure 5.13 Distributions of side chain virtual bond angles from MD simulations of dipeptides using the GBSW and HyRes models.

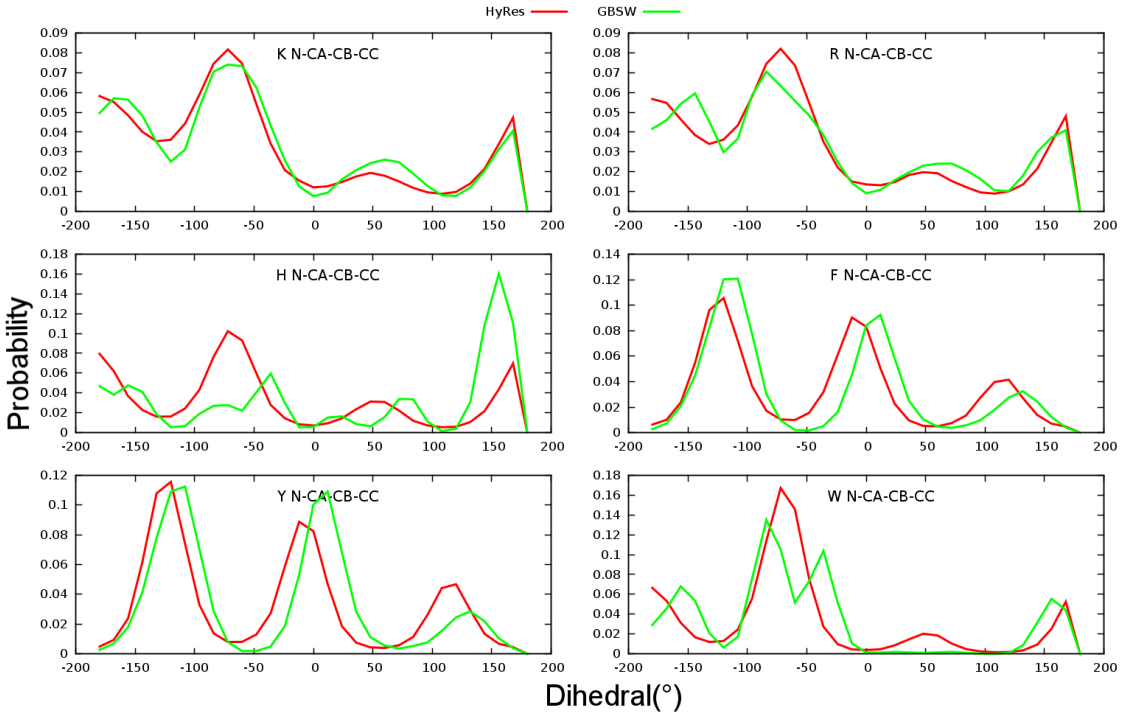


Figure 5.14 Probability distributions of dihedral χ in dipeptides obtained from CG and GBSW atomistic simulations.

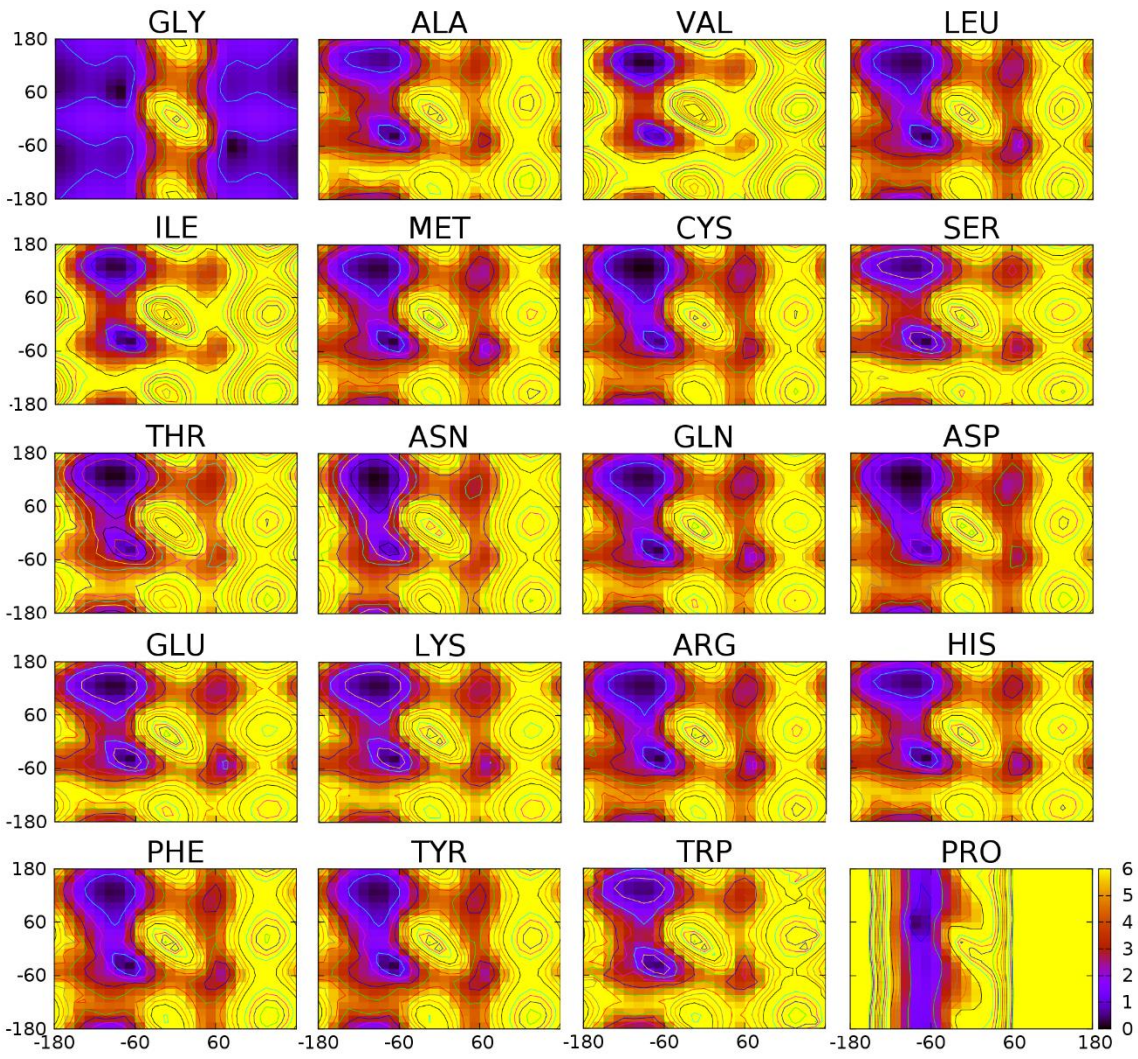


Figure 5.15 Backbone ϕ/ψ adiabatic energy surfaces (in kcal/mol) of all 20 dipeptides in the HyRes model. The surface was calculated by energy minimization with ϕ and ψ restrained at specified values. The surface was shifted such that the minimum value for each system was zero.

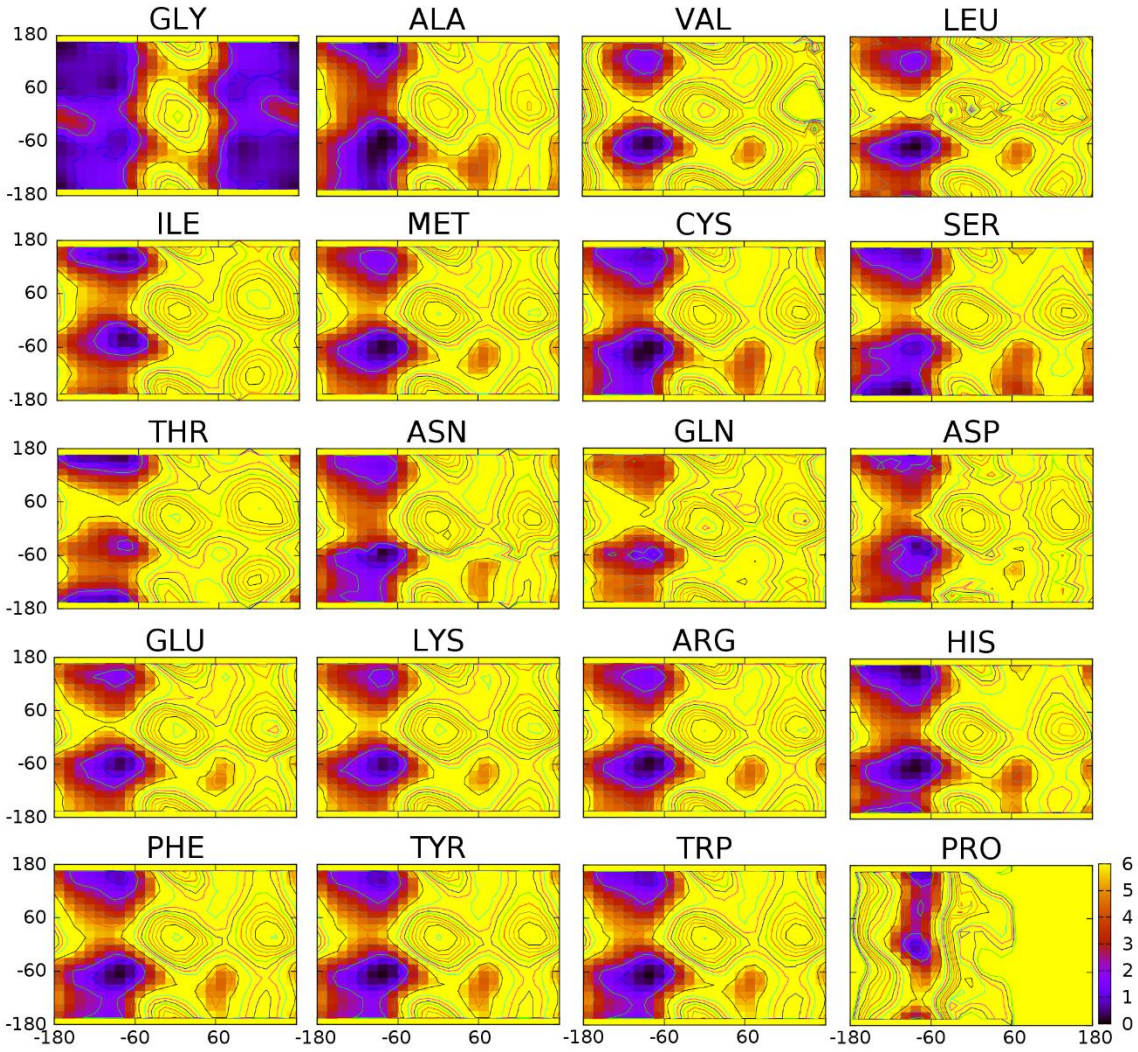


Figure 5.16 Backbone ϕ/ψ adiabatic energy surfaces (in kcal/mol) of all 20 dipeptides in the GBSW implicit solvent model. The surface was calculated by energy minimization with ϕ and ψ restrained at specified values. The surface was shifted such that the minimum value for each system was zero.

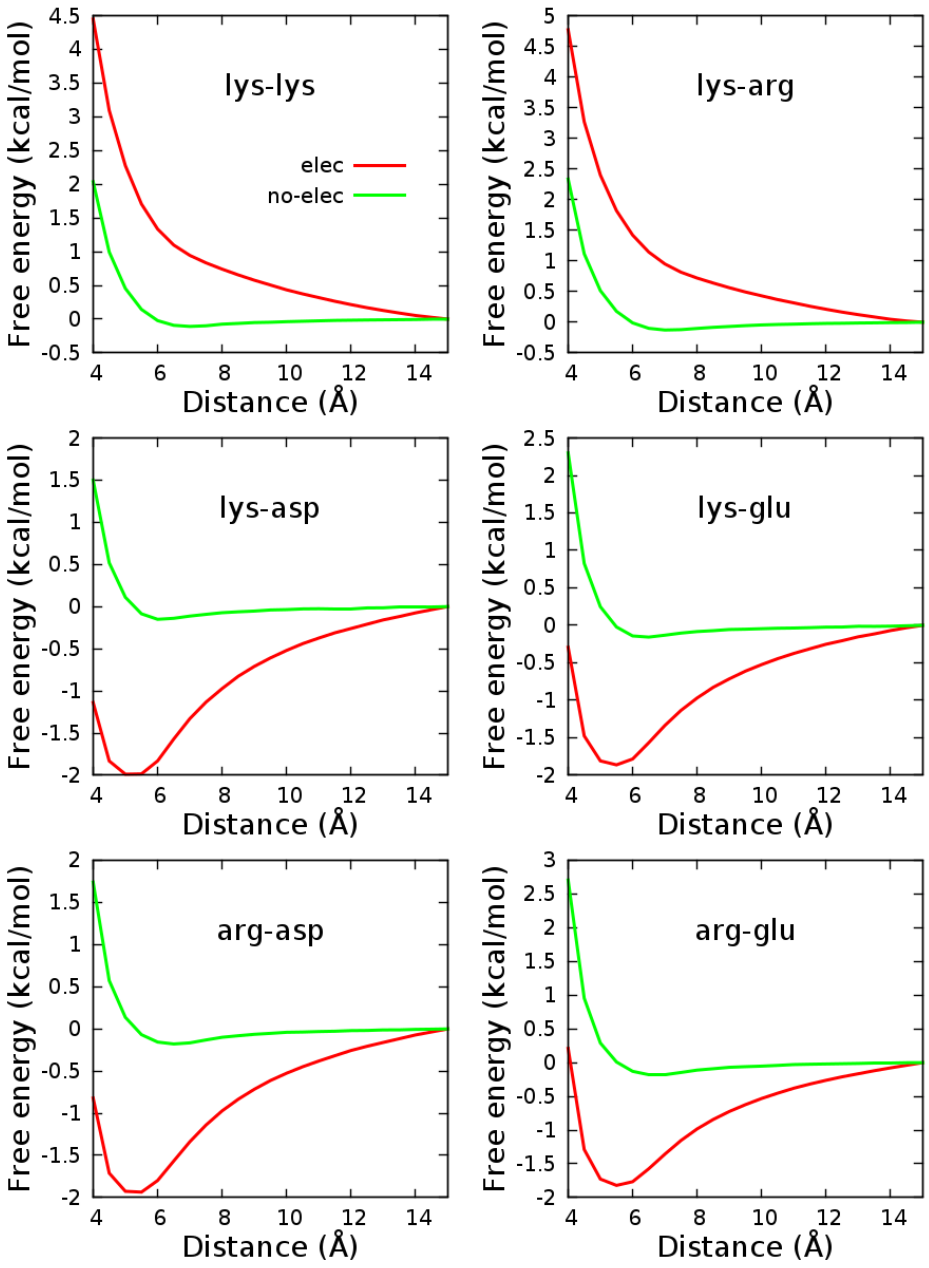


Figure 5.17 Free energy profiles as a function of separation distance between the center of mass of charged amino acid side chain analogs with (red traces) and without (green traces) electrostatic interactions.

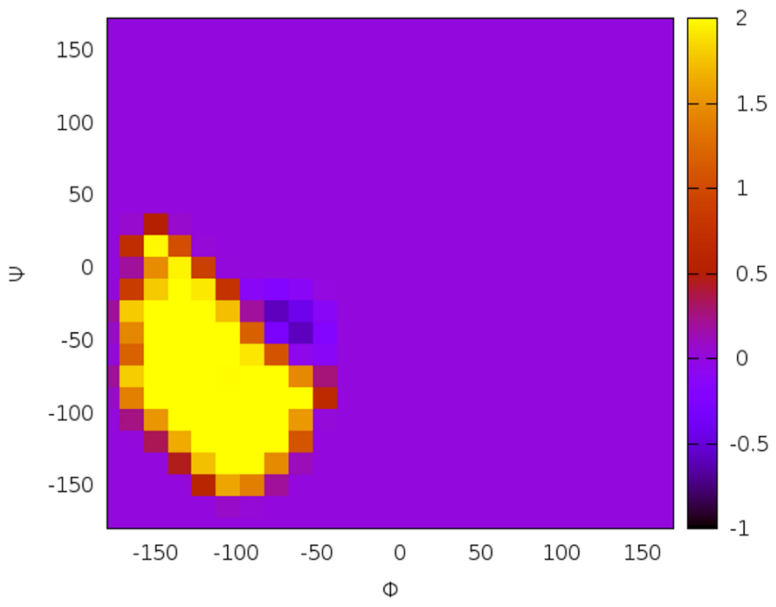


Figure 5.18 Backbone φ/ψ CMAP cross-term (in kcal/mol) in the HyRes model, which includes a small energy basin to stabilize α-helixes and a energy barrier to suppress the sampling of π-helixes.

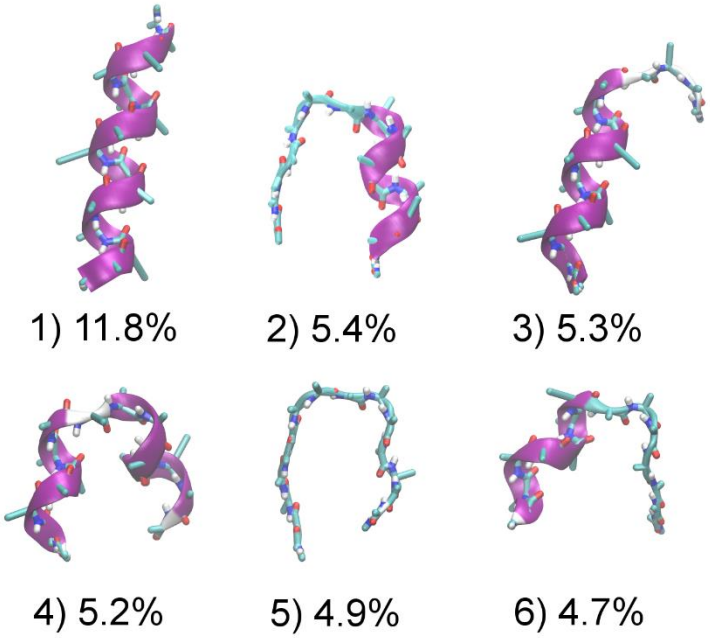


Figure 5.19 Centroids and populations of six largest clusters for $(AAQAA)_3$ structure ensemble in the control simulation using HyRes model. All heavy atoms were used to compute RMSD between structures, and a fixed radius of 4 Å was used to define clusters.

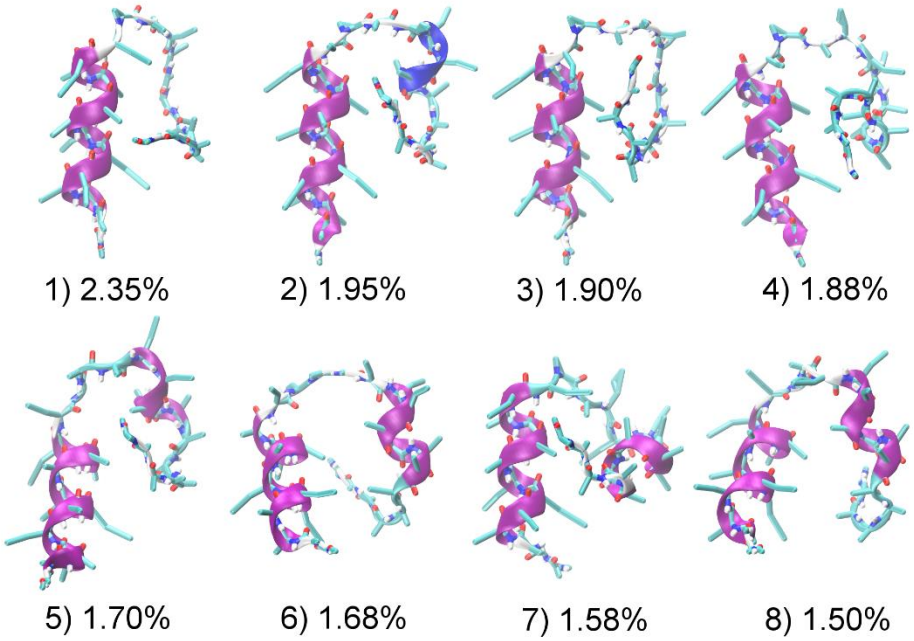


Figure 5.20 Centroids and populations of eight largest clusters for KID structure ensemble in the control simulation using HyRes model. CA and CB atoms were used to compute RMSD between structures, and a fixed radius of 4 Å was used to define clusters.

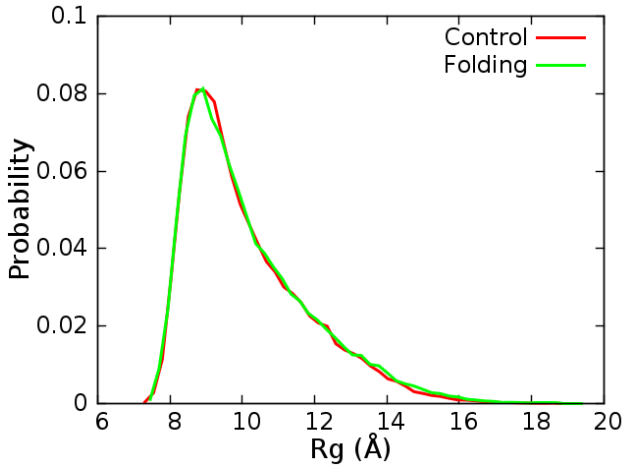


Figure 5.21 Probability distribution of R_g of RS peptide obtained from HyRes simulations at 300 K.

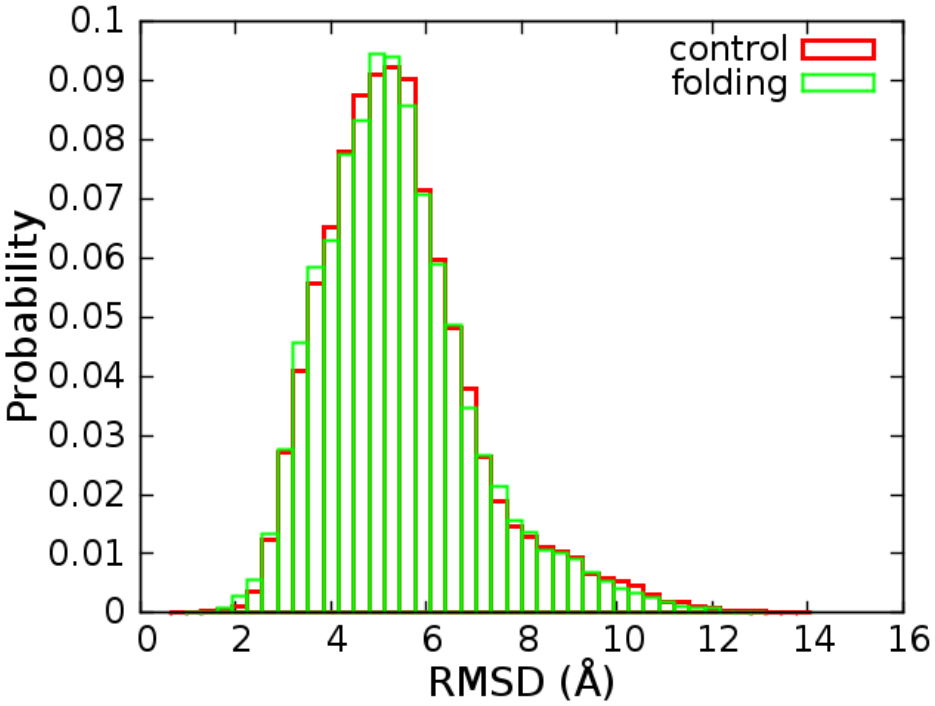


Figure 5.22 Probability distribution of backbone RMSD of GB1m3 peptide with respect to the folded state from HyRes simulations at 300 K.

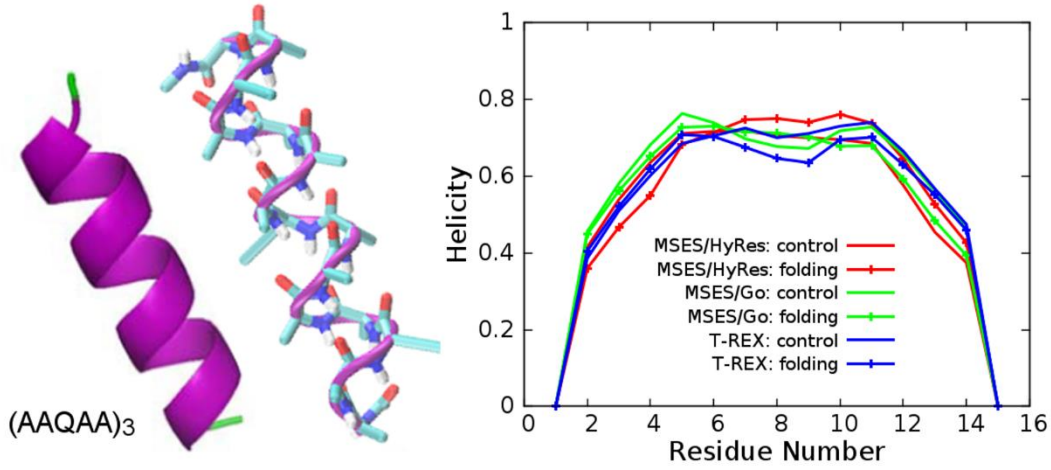


Figure 5.23 (Left) AT and HyRes representation of folded $(AAQAA)_3$. The AT model is shown in cartoon and HyRes model in Licorice representation, with the backbone highlighted in purple. (Right) Residual helicity profiles of $(AAQAA)_3$ at 270 K obtained from MSES/HyRes, MSES/Gō, and T-REX simulations (Table 5.1 of the main text). Note that only the second half of each trajectory was used in these calculations.

CHAPTER 6

MULTISCALE ENHANCED SAMPLING WITH SOLUTE TEMPERING FOR STUDYING DISORDERED PROTEIN CONFORMATIONS

Intrinsically disordered proteins (IDPs) are fully functional but don't have well-defined structures under physiological conditions. The highly dynamic and heterogeneous structural ensembles of IDPs are not amenable to traditional experimental characterizations, and searching their vast conformational space also pushes the limit of sampling capacity of computational simulations. Recently it has been realized that multiscale enhanced sampling (MSES) method may be highly suitable for accelerating atomistic simulations of IDPs by coupling the atomistic model to coarse-grained (CG) ones. Bias from coupling to CG models can be removed using Hamiltonian replica exchange such that one could benefit simultaneously from the high accuracy of atomistic models and fast dynamics of CG ones. To maximize sampling efficiency of MSES simulations, structural transitions at the CG level should be as fast as possible and properly translated to the atomistic level. Also, the replica exchange protocol should allow the coupled conditions to be rapidly exchanged to the uncoupled/unbiased limit. In the present study, we presented a simple but effective strategy to achieve these goals, by integrating MSES with solute tempering (MSES-ST). Here, the effective temperature of CG model in all conditions is set to its melting temperature using the Hamiltonian scaling approach based on the idea of solute tempering. This will maximize the structural transition rate at the CG level, thus further accelerating atomistic structural transitions. Meanwhile, replica mixing could be sped up. Application of MSES-ST

to calculate the conformational equilibria of a non-trivial IDP demonstrates its improved sampling efficiency over the original MSES scheme.

6.1 Introduction

Being fully or partially disordered under physiological conditions, intrinsically disordered proteins (IDPs) are recently recognized functional proteins that deviate from the traditional structure-function paradigm[1-4, 6, 7, 17, 21, 169, 178]. About one-third of eukaryotic proteins contain disordered regions with lengths of more than 40 amino acids[13], which implies that many proteins rely on intrinsic disorder to carry out functions. Indeed, there is mounting evidence that IDPs could interact with many targets and often play critical roles in cellular signaling and regulation[3, 13, 17, 21, 107, 170]. Moreover, missense mutation of disordered regions is frequently associated with human diseases, like cancer and neuron degenerative diseases[13, 170, 173, 174, 240]. Thus, there is a great need to understand the molecular mechanisms of the functional roles of IDPs in signal transduction and regulation network.

However, the highly dynamic and heterogeneous conformations of IDPs pose great challenges in their experimental measurements, from which only ensemble-averaged properties can be obtained in most cases[55, 166, 206, 241, 304]. In contrast, molecular dynamics simulations using physics-based atomistic models could provide high temporal and spatial resolution information of the system of interest, thus offering a powerful tool in characterizing IDPs. Such approach, nevertheless, is computationally costly and often difficult to generate converged

structural ensembles for IDPs. Therefore, enhanced sampling method usually is required to reduce the computational cost and accelerate atomistic simulations of IDPs.

Development of enhanced sampling technique has been under intensive research, and many methods have been proposed, among which temperature replica exchange (T-RE)[78] is one of the most popular protocols. Unfortunately, the activation free energy of large conformational rearrangement, like protein folding, is often dominated by the entropic component[93, 244], which makes tempering ineffective in accelerating such transitions[305, 306]. To overcome such entropic problem, we have recently developed an enhanced sampling technique, called multiscale enhanced sampling (MSES) method by taking advantage of the reduced conformational space of coarse-grained (CG) modeling and coupling CG models with atomistic (AT) ones to accelerate sampling of atomistic protein energy landscapes [83, 243]. Briefly, the simulation box contains both AT and CG representations of the system. The AT and CG components of the hybrid system don't interact with each other directly, but configurations at both resolutions are coupled through a MSES coupling potential (U_{MSES}) that is applied to selected degrees of freedom:

$$U_{hybrid}(r_{AT}, r_{CG}, \lambda) = U_{AT}(r_{AT}) + U_{CG}(r_{CG}) + \lambda U_{MSES}(r_{AT}, r_{CG}) \quad (6.1)$$

where U_{hybrid} is the potential energy of the hybrid system, and U_{AT} and U_{CG} are potentials for AT and CG models, respectively. The MSES coupling potential, U_{MSES}, is to restrain the structural difference between AT and CG representations along selected degrees of freedom. Such coupling scheme makes MSES more tolerant to

inevitable artifacts in CG models and scalable to large systems. U_{MSES} is further scaled by a coupling factor λ that ranges from 0 to 1. When λ is large, the AT and CG models are strongly coupled, and faster structural transitions at the CG level are expected to drive AT conformational transitions. To unbias any artifacts generated by coupling AT and CG models, Hamiltonian replica exchange is utilized, such that proper ensembles could be recovered at both AT and CG levels at the limit condition of $\lambda = 0$. Meanwhile, T-RE is incorporated to further enhance the sampling efficiency.

Previous MSES simulations of both partial helical peptides[243] and β-hairpin[83] have shown that this strategy is able to significantly improve sampling efficiency, but these studies also suggest that maximizing MSES efficiency requires advanced protocols to further accelerate structural transitions at AT (and CG) levels in strongly coupled conditions. Also, such coupled conditions should be rapidly exchanged to the uncoupled limit so as to expedite convergence of simulated ensembles. In the present work, we present a simple but effective strategy to achieve these goals, by integrating MSES with solute tempering[183] (MSES-ST). In replica exchange with solute tempering, both solute-solute and solute-solvent interactions are scaled down, such that the effective temperature of solute can be increased. Here, similar Hamiltonian scaling approach is used to set the effective temperature of CG model in all conditions at its melting temperature. Since AT and CG models don't directly interact with each other, only CG potential needs to be scaled. We show that this strategy is effective in accelerating atomistic conformational rearrangements as well

as increasing replica mixing rate, which facilitates rapid generation of converged structural ensembles of IDPs.

6.2 Methods

6.2.1 Simulation details

A β-hairpin peptide GB1p (GEWTY DDATK TFTVT E), was used to evaluate the sampling efficiency of MSES-ST. As a comparison, simulations using T-RE and original MSES were also performed. All T-RE, MSES and MSES-ST simulations were carried out using CHARMM [149, 150] together with modified MMTSB[146]. For each protocol, two independent simulations were performed, one starting from fully folded state (i.e., control run) and the other one from fully extended state (i.e., folding run). Each simulation had eight replicas, with their temperatures spaced exponentially between 300 K and 450 K. Hamiltonian of the CG model in MSES-ST simulations was scaled such that its effective temperature in all replicas was 389 K. The atomistic model used in all simulations was implicit solvent GBSW force field[59], while in MSES and MSES-ST simulations, the atomistic model was also coupled with a topology-based coarse-grained model, the Gō-like model calibrated in our previous work[83].

In MSES and MSES-ST simulations, the coupling potential was applied to Cα-Cα distances between nine residue pairs that form native contacts. To reduce energy penalty for large structural deviations between atomistic and CG copies and ensure uniform exchange acceptance probability between neighboring replicas, the coupling potential, U_{MSES}, was smoothly switched from a quadratic form for small structural deviations to the soft asymptote for large deviations[243]:

$$U_{MSES}(r_{AT}, r_{CG}) = \sum_i 0.5k_i\left(d_i^{AT} - d_i^{CG}\right)^2, \qquad if\ |\Delta d_i| \le d_s$$

$$= \sum_i A + \frac{B}{\left(d_i^{AT} - d_i^{CG}\right)^s} + f_{max}\left(d_i^{AT} - d_i^{CG}\right), \qquad if\ |\Delta d_i| > d_s$$

$$with\ \Delta d_i = d_i^{AT} - d_i^{CG} \tag{6.2}$$

where d_i^{AT} and d_i^{CG} are Cα-Cα distances for the *i*-th native contact in the atomistic and CG models, respectively, and Δd_i is the difference between them. At the distance threshold d_s, U_{MSES} begins to smoothly switch from harmonic form to the soft asymptote. The switching exponent *s* indicates how fast the limiting force, f_{max}, can be approached at large $|\Delta d_i|$. Parameters *A* and *B* are calculated by requiring both MSES coupling energy and force to be continuous when $|\Delta d_i|$ is at the threshold distance d_s. Parameters used in the present work were k = 1.0 kcal/mol/Å^2, s = 1, d_s = 2.0 Å, and f_{max} = 0.1 kcal/mol/Å. The coupling potential was further scaled by a coupling parameter λ (see **Eq. 6.1**) ranging from 0 to 1, and λ values of the eight replicas were 0, 0.10, 0.22, 0.35, 0.49, 0.64, 0.81 and 1.00, respectively.

For all simulations, Langevin dynamics with a friction coefficient of 0.1 ps^{-1} was performed. The equation of motion was integrated with a time step of 2 fs. SHAKE algorithm [156] was used to constrain the length of all bonds involving hydrogen atoms. Exchange of replicas with their neighbors were attempted every 2 ps. The production simulation time was 800 ns/replica in all case.

6.2.2 Analysis

To compute the number of native hydrogen bonds in the atomistic model of GB1p during simulations, we first computed the distances between seven pairs of

backbone N and O atoms that are involved in hydrogen bonding in the fully folded state. In comparison to the fully folded state, the seven distance violations were calculated, which then can be used to compute the number of native hydrogen bonds as follows:

$$N = \sum_{i=1}^{7} 1 - \frac{1}{1 + e^{-4\left(\Delta d_i - \frac{7}{4}\right)}} \tag{6.3}$$

where Δd_i is the distance violation for the i-th pair of N-O atoms, and N is the number of native hydrogen bonds.

The atomistic structural transition rate was calculated as the total number of reversible folding-unfolding transitions divided by the total simulation time, where folded and unfolded states were defined as $N \geq 6.5$ and $N \leq 0.5$, respectively. Results reported in this study are averaged values over all replicas.

To compute simulation error as a function of trajectory length t, each trajectory was first divided into multiple segments of length t. For each segment, the distribution of native hydrogen bond number was calculated, and mean absolute error was computed with respect to the reference profile, which was the distribution of native hydrogen bond number calculated using the whole trajectory and averaged over all six simulations. Average and standard deviation of mean absolute error were reported here by considering all trajectory segments of length t.

6.3 Results and Discussion

By design, both MSES and its variant MSES-ST could generate correct ensembles at the condition of $\lambda = 0$ via Hamiltonian replica exchange. As shown in Figure 6.1,

the probability distributions of number of native hydrogen bonds in the atomistic model obtained from T-RE, MSES and MSES-ST simulations are converged to the same profile, suggesting that no thermodynamic bias is introduced in MSES and MSES-ST protocols.

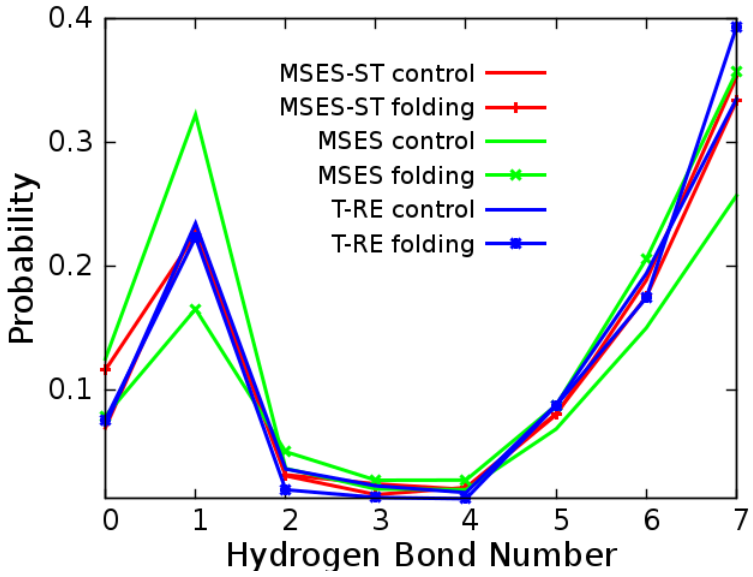


Figure 6.1 Probability distributions of number of native hydrogen bonds in AT model at 300 K obtained from T-RE, MSES, and MSES-ST simulations, including both control and folding runs.

Figure 6.2 demonstrates the reversible folding-unfolding transition rate in the atomistic GBSW model. Clearly, coupling the atomistic model to even a simple topology-based CG model in both MSES and MSES-ST simulations could accelerate structural transitions at the atomistic level. Moreover, setting the effective temperature of CG model to its melting temperature in MSES-ST simulations appeared to further facilitate atomistic structural transitions. Note that the enhancement of structural transition rate in MSES-ST depends on the system of interest and the CG model used. We believe that further improvement of the quality of CG modeling may lead to even faster conformational rearrangement. In any case,

considering that the sampling efficiency of T-RE is often hindered by slow structural transitions, such rapid structural transitions in MSES-ST should be beneficial for enhanced sampling capacities and generation of well-converged ensembles.

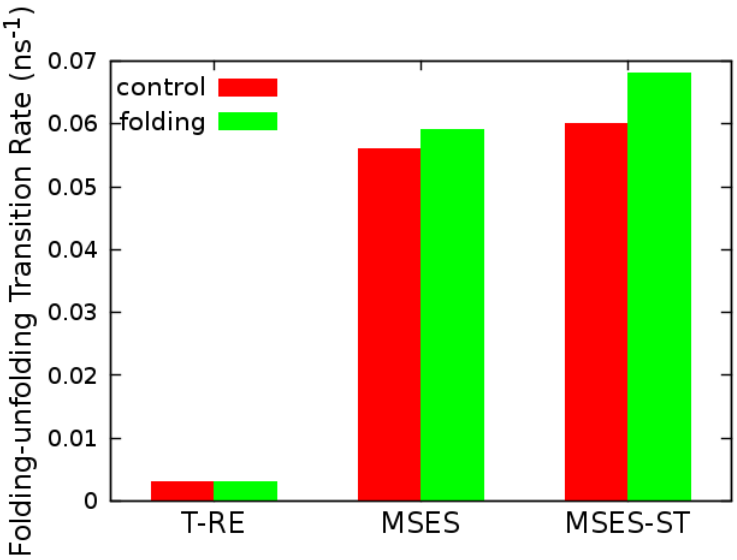


Figure 6.2. Reversible folding-unfolding transition rate in the atomistic model of GB1p averaged over all replicas.

Another indicator of sampling efficiency is the rate of replica round trips between the lowest and highest temperatures, or simply replica mixing rate. There have been many efforts towards maximizing this quantity to further increase the sampling efficiency of T-RE, including adaptively adjusting the distribution of simulation temperatures[84, 85] or using biasing potentials[86, 87]. Intriguingly, the replica mixing rate in MSES-ST is the fastest among these three simulation protocols (see Figure 6.3). Such rapid replica mixing allows the fast dynamics driven by CG models under the coupled conditions to be promptly exchanged to uncoupled limit, thus increasing sampling efficiency.

Due to the fast replica mixing, as well as rapid atomistic conformational transitions driven by a CG model in its melting temperature, the sampling efficiency of MSES-

ST is the highest among the three protocols tested here, as shown in Figure 6.4. The mean absolute error in calculating the probability distribution of number of native hydrogen bonds in GB1p decays fastest in MSES-ST simulations, suggesting its superior sampling capacity.

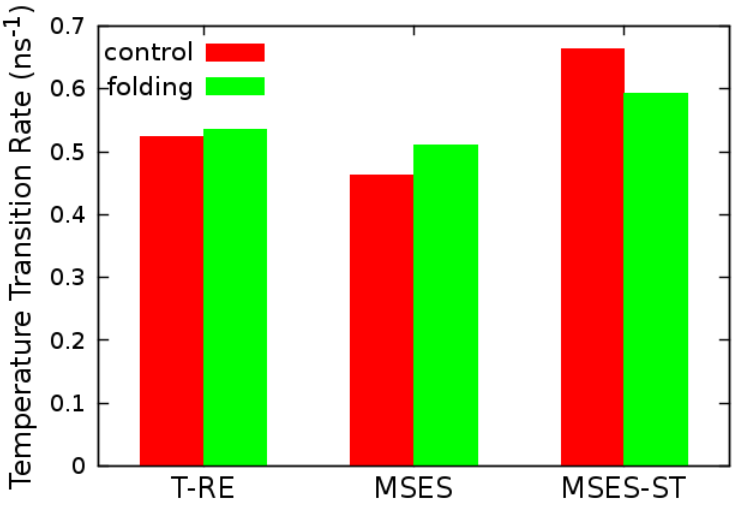


Figure 6.3. Rate of replica round trips between the lowest and highest temperatures averaged over all replicas.

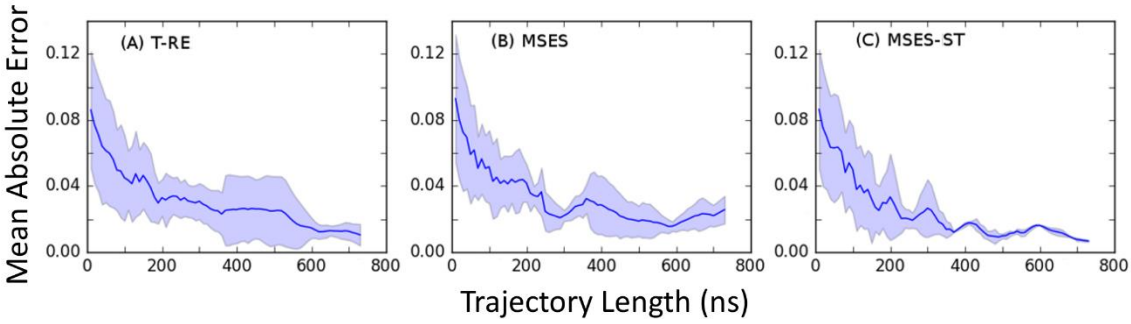


Figure 6.4. Mean absolute error as a function of simulation trajectory length in calculating the probability distribution of native hydrogen bond number of GB1p at 300 K. The solid lines are for averaged values over multiple trajectories, and the shaded regions indicate the standard deviation. See section Analysis for more details.

6.4 Conclusions

Understanding the structure and dynamics of IDPs plays key roles in elucidating the molecular mechanisms of how IDPs carry out versatile functions. It has been recently recognized that MSES may be suitable for atomistic characterization of the highly dynamic, heterogeneous structural ensembles of IDPs, although maximizing the sampling efficiency requires the CG model to drive atomistic structural transitions as fast as possible and the coupled conditions to be rapidly exchanged to the uncoupled/unbiased condition. To achieve this goal, here we are presenting a simple but effective strategy by integrating MSES with solute tempering (MSES-ST). According to the idea of solute tempering, the effective temperature of CG model in all conditions can be set to its melting temperature by scaling its Hamiltonian. This will maximize the structural transition rate at the CG level, and further accelerate atomistic structural transitions. Such design also increases replica mixing rate. Application of MSES-ST to a model IDP demonstrates that faster atomistic structural transitions can be achieved, which leads to rapider generation of converged structural ensembles and demonstrates its improved sampling efficiency over the original MSES scheme.

CHAPTER 7

SUMMARY AND FUTURE DIRECTIONS

7.1 Summary

It has been well recognized that IDPs are important functional proteins in many cellular processes, especially cellular signaling and regulation, but the mechanisms of how intrinsic disorder mediates protein functions remain elusive. In this dissertation, computer simulations have been used to study the structure, dynamics and interactions of IDPs, which provides important mechanistic insights into the functional roles of IDPs. In Chapter 2 and 3, atomistic simulations have been used to study the mechanisms of regulated unfolding of Bcl-xL induced by PUMA binding. By characterizing the structural ensemble of unbound Bcl-xL, we found that the PUMA-binding interface of Bcl-xL is intrinsically disordered and could sample various previously observed conformations, including both unbound and bound states. Such intrinsic disorder appears to provide the physical basis of Bcl-xL regulated unfolding upon interacting with PUMA, and maybe a general mechanism of how Bcl-xL could respond sensitively and rapidly to all sorts of cellular signals. Chapter 4 presents a mechanistic study of how pre-formed helical element in an IDP, ACTR, modulates the kinetics of its binding to another IDP, NCBD. By performing simulations using a series of carefully calibrated coarse-grained models, it has been found that residual structure in ACTR accelerates NCBD binding mainly by promoting efficient folding upon encounter. These studies provide important mechanistic insights into how IDPs carry out functions in signal transduction and

regulation network, but at the same time, reveal the challenges in computational modelling and simulations in studying IDPs, especially the urgent need of new sampling techniques for more efficient atomistic simulations.

Our group has recently developed a novel enhanced sampling method, multiscale enhanced sampling (MSES), to improve the efficiency of atomistic simulations. In MSES, a coarse-grained model is coupled to the atomistic one, such that atomistic structural transitions could be sped up due to the fast dynamics of coarse-grained models. Any bias introduced from such coupling can be removed using Hamiltonian replica exchange. Thus, MSES allows us to benefit simultaneously from the high accuracy of atomistic models and fast dynamics of coarse-grained models. To further improve the efficiency of MSES protocol, two strategies have been demonstrated in this dissertation. Chapter 5 presents a new hybrid-resolution coarse-grained model for IDPs, named HyRes. By representing the protein backbone at the atomistic level and side chains at the coarse-grained level and optimizing interaction parameters targeting results from atomistic simulations, experimental studies, HyRes could semi-quantitatively describe the secondary structure of IDPs and qualitatively describe the long-range interactions. This allows HyRes to generate fluctuations largely consistent with those in atomistic models, thus improving the sampling efficiency of MSES. In chapter 6, an advanced replica exchange protocol is shown, named MSES with solute tempering (MSES-ST). Here, the effective temperature of coarse-grained model in MSES is set at its melting temperature by scaling the Hamiltonian. This design allows coarse-grained model to generate faster structural transitions, thus further expediting conformational arrangement at the

atomistic level. Replica mixing rate is also increased, which enables the fast dynamics in coupled conditions to be rapidly communicated to the uncoupled limit. Therefore, MSES-ST could show superior sampling capacity than the original MSES protocol.

7.2 Future directions

The traditional protein structure-function paradigm had long been viewed as one of the central scientific dogmas in molecular biology, which suggests that functions of protein are tightly linked to their well-defined 3D structures. However, the discovery of IDPs has greatly challenged this notion. Although lack of well-defined 3D structures in physiological conditions, IDPs are highly abundant and functionally important in biological systems. It has now been recognized that protein functions can also be mediated by conformational disorder and timescale of motions. Thus, there is a great need to establish a new, unified understanding of protein structure/disorder/dynamics-function relationship.

Understanding such new “structure”-function relationship will also help us better understand how pathological conditions, like mutations or abnormal changes in cellular environment, may impact protein structure, dynamics, interactions and functions in human diseases. Moreover, such mechanistic understanding will lay a critical foundation for future development of therapeutic strategies to treat these diseases. For instance, an emerging approach in drug design is to target disease-associated IDPs, by disrupting interactions of IDPs with their targets, thus altering their functions. Drug molecules could compete with IDPs for binding to the target.

Alternatively, they could directly interact with IDPs and change the structure and/or dynamics of IDPs, thus preventing the association of IDPs with their targets. These novel strategies will greatly expand the venue of drug design, although their successful realization will depend critically on our ability to obtain high resolution information of the structure, dynamics and interactions of proteins.

In this regard, computational simulation has proven a powerful tool, especially in characterizing the highly heterogeneous and dynamic structures of IDPs. Atomistic simulations in particular could provide the necessary high-resolution temporal and spatial information that is otherwise unavailable. Of course, challenges remain in computational simulation of IDPs using atomistic models, especially the accuracy of force fields and the efficiency of sampling algorithm. To improve the accuracy of current force fields, it requires both correct identification of sources of error and systematic reparameterization accordingly. In order to achieve this goal, a large amount of IDPs (and folded proteins) need to be simulated, and the generated ensembles should be compared with experimental measurements. Considering the computational cost of atomistic simulations, it remains difficult to obtain fully converged structural ensembles for IDPs, which then hinders the improvement of force field development. Therefore, developing enhanced sampling methods to accelerate atomistic simulations is of great importance.

The challenges in sampling the conformational space of IDPs have multiple aspects. As discussed in this dissertation, the free energy barriers of larger, cooperative structural transitions are often dominated by the entropic component, which makes many sampling methods that are based on tempering ineffective. Therefore, coarse-

grained modelling would be a good alternative strategy to overcome entropic barrier and drive atomistic structural transitions due to the reduced resolution. However, how to maximize the sampling efficiency of such approaches needs further exploration. Another challenge in conformational sampling is that hidden barrier in the orthogonal space often exists for nontrivial structural transitions. For instance, certain order parameter can be defined to describe the protein folding process, but direct move of this order parameter using biasing potentials may not necessarily activate protein folding, since environment also needs to respond properly, like the rearrangement of solvent molecules. In this case, orthogonal space sampling could be used to enhance sampling along both selected order parameter and its environment response, by adding biasing potentials as a function of the order parameter and its generalized force. Unfortunately, there is often no obvious order parameter to describe the protein conformational rearrangement, especially in the case of IDPs where numerous sub-states may exist and there might be multiple pathways connecting two macroscopic states. Since MSES scheme doesn't require any definition of order parameters, we propose that integration of MSES with orthogonal space sampling in the future may help alleviate the problems of entropic barrier, hidden barrier from slow environmental response, and lack of well-defined order parameters. With advanced sampling techniques, more efficient atomistic simulations can be performed to characterize IDPs in details, which could help us better understand the new protein "structure"-function relationship and design more effective therapeutic strategies to treat human diseases.

BIBLIOGRAPHY

[1] Wright PE, Dyson HJ. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J Mol Biol. 1999;293:321-31.
[2] Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Bio. 2005;6:197-208.
[3] Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recognit. 2005;18:343-84.
[4] Dunker AK, Silman I, Uversky VN, Sussman JL. Function and structure of inherently disordered proteins. Curr Opin Struc Biol. 2008;18:756-64.
[5] Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. J Mol Graph Model. 2001;19:26-59.
[6] Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci. 2002;27:527-33.
[7] Click TH, Ganguly D, Chen J. Intrinsically disordered proteins in a physics-based world. Int J Mol Sci. 2010;11:5292-309.
[8] Boesch C, Bundi A, Oppliger M, Wuthrich K. 1H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. European journal of biochemistry. 1978;91:209-14.
[9] Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, et al. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing. 1998:473-84.
[10] Garner E, Cannon P, Romero P, Obradovic Z, Dunker AK. Predicting Disordered Regions from Amino Acid Sequence: Common Themes Despite Differing Structural Characterization. Genome informatics Workshop on Genome Informatics. 1998;9:201-13.
[11] Romero P, Obradovic Z, Li XH, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins-Structure Function and Genetics. 2001;42:38-48.
[12] Vucetic S, Brown CJ, Dunker AK, Obradovic Z. Flavors of protein disorder. Proteins-Structure Function and Genetics. 2003;52:573-84.
[13] Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol. 2002;323:573-84.
[14] Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: Introducing the D-2 concept. Annual review of biophysics. 2008;37:215-46.
[15] Hilser VJ, Thompson EB. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. Proc Natl Acad Sci U S A. 2007;104:8311-5.
[16] Chen JH. Atomistic Glimpse of the Orderly Chaos of One Protein. Biophys J. 2015;109:1511-2.
[17] Smock RG, Gierasch LM. Sending Signals Dynamically. Science. 2009;324:198-203.
[18] Bah A, Forman-Kay JD. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. J Biol Chem. 2016;291:6696-705.
[19] Csizmok V, Follis AV, Kriwacki RW, Forman-Kay JD. Dynamic Protein Interaction Networks and New Structural Paradigms in Signaling. Chemical reviews. 2016;116:6424-62.

[20] Shammas SL, Rogers JM, Hill SA, Clarke J. Slow, reversible, coupled folding and binding of the spectrin tetramerization domain. Biophys J. 2012;103:2203-14.
[21] Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Bio. 2015;16:18-29.
[22] Shin Y, Brangwynne CP. Liquid phase condensation in cell physiology and disease. Science. 2017;357.
[23] Berlow RB, Dyson HJ, Wright PE. Expanding the Paradigm: Intrinsically Disordered Proteins and Allosteric Regulation. J Mol Biol. 2018;430:2309-20.
[24] Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proc Natl Acad Sci U S A. 2000;97:8868-+.
[25] Fuxreiter M, Simon I, Friedrich P, Tompa P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol. 2004;338:1015-26.
[26] Kjaergaard M, Teilum K, Poulsen FM. Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. Proc Natl Acad Sci U S A. 2010;107:12535-40.
[27] Uversky VN. Intrinsically Disordered Proteins and Their Environment: Effects of Strong Denaturants, Temperature, pH, Counter Ions, Membranes, Binding Partners, Osmolytes, and Macromolecular Crowding. Protein J. 2009;28:305-25.
[28] Dyson HJ, Wright PE. Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. Method Enzymol. 2001;339:258-70.
[29] Dyson HJ, Wright PE. Unfolded proteins and protein folding studied by NMR. Chemical reviews. 2004;104:3607-22.
[30] Chemes LB, Alonso LG, Noval MG, de Prat-Gay G. Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains. Methods Mol Biol. 2012;895:387-404.
[31] Iešmantavičius V, Dogan J, Jemth P, Teilum K, Kjaergaard M. Helical propensity in an intrinsically disordered protein accelerates ligand binding. Angewandte Chemie International Edition. 2014;53:1548-51.
[32] LeBlanc SJ, Kulkarni P, Weninger KR. Single Molecule FRET: A Powerful Tool to Study Intrinsically Disordered Proteins. Biomolecules. 2018;8.
[33] Mukhopadhyay S, Krishnan R, Lemke EA, Lindquist S, Deniz AA. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. Proc Natl Acad Sci U S A. 2007;104:2649-54.
[34] Bernado P, Svergun DI. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. Molecular BioSystems. 2012;8:151-67.
[35] Gabel F. Small angle neutron scattering for the structural study of intrinsically disordered proteins in solution: a practical guide. Methods Mol Biol. 2012;896:123-35.
[36] Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural characterization of flexible proteins using small-angle X-ray scattering. Journal of the American Chemical Society. 2007;129:5656-64.
[37] Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci U S A. 2005;102:17002-7.

[38] Huang F, Rajagopalan S, Settanni G, Marsh RJ, Armoogum DA, Nicolaou N, et al. Multiple conformations of full-length p53 detected with single-molecule fluorescence resonance energy transfer. Proceedings of the National Academy of sciences. 2009;106:20758-63.
[39] Lee H, Mok KH, Muhandiram R, Park KH, Suk JE, Kim DH, et al. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. J Biol Chem. 2000;275:29426-32.
[40] Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, et al. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci U S A. 2008;105:5762-7.
[41] Trexler AJ, Rhoades E. Single Molecule Characterization of alpha-Synuclein in Aggregation-Prone States. Biophys J. 2010;99:3048-55.
[42] Marsh JA, Singh VK, Jia ZC, Forman-Kay JD. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: Implications for fibrillation. Protein Science. 2006;15:2795-804.
[43] Bernado P, Bertoncini CW, Griesinger C, Zweckstetter M, Blackledge M. Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings. Journal of the American Chemical Society. 2005;127:17968-9.
[44] Zagorski MG, Hou LM, Shao HY, Zhang YB, Menon N. Solution nmr studies of the AB(1-40) and AB(1-42) peptides establish that the Met35 oxidation state affects the mechanism of amyloid formation. Neurobiol Aging. 2004;25:S153-S.
[45] Yan YL, McCallum SA, Wang CY. M35 oxidation induces A beta 40-like structural and dynamical changes in A beta 42. Journal of the American Chemical Society. 2008;130:5394-+.
[46] Roche J, Shen Y, Lee JH, Ying JF, Bax A. Monomeric A beta(1-40) and A beta(1-42) Peptides in Solution Adopt Very Similar Ramachandran Map Distributions That Closely Resemble Random Coil. Biochemistry. 2016;55:762-75.
[47] Granata D, Baftizadeh F, Habchi J, Galvagnion C, De Simone A, Camilloni C, et al. The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. Sci Rep-Uk. 2015;5.
[48] Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. Nature chemical biology. 2009;5:789-96.
[49] Fuxreiter M. Fold or not to fold upon binding—does it really matter? Curr Opin Struc Biol. 2019;54:19-25.
[50] Delaforge E, Kragelj J, Tengo L, Palencia A, Milles S, Bouvignies G, et al. Deciphering the Dynamic Interaction Profile of an Intrinsically Disordered Protein by NMR Exchange Spectroscopy. Journal of the American Chemical Society. 2018;140:1148-58.
[51] Milles S, Mercadante D, Aramburu IV, Jensen MR, Banterle N, Koehler C, et al. Plasticity of an Ultrafast Interaction between Nucleoporins and Nuclear Transport Receptors. Cell. 2015;163:734-45.
[52] Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature. 2007;447:1021-5.
[53] Chen JH. Towards the physical basis of how intrinsic disorder mediates protein function. Archives of biochemistry and biophysics. 2012;524:123-31.

[54] Haq SR, Chi CN, Bach A, Dogan J, Engstrom A, Hultqvist G, et al. Side-Chain Interactions Form Late and Cooperatively in the Binding Reaction between Disordered Peptides and PDZ Domains. Journal of the American Chemical Society. 2012;134:599-605.
[55] Schuler B, Eaton WA. Protein folding studied by single-molecule FRET. Curr Opin Struc Biol. 2008;18:16-26.
[56] Beberg AL, Ensign DL, Jayachandran G, Khaliq S, Pande VS. Folding@home: Lessons From Eight Years of Volunteer Distributed Computing. Int Parall Distrib P. 2009:1624-+.
[57] Shirts M, Pande VS. Computing - Screen savers of the world unite! Science. 2000;290:1903-4.
[58] Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, et al. Anton, a special-purpose machine for molecular dynamics simulation. Commun Acm. 2008;51:91-7.
[59] Chen JH, Im WP, Brooks CL. Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. Journal of the American Chemical Society. 2006;128:3728-36.
[60] Vitalis A, Pappu RV. ABSINTH: A New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions. J Comput Chem. 2009;30:673-99.
[61] Ganguly D, Chen J. Atomistic details of the disordered states of KID and pKID. implications in coupled binding and folding. J Am Chem Soc. 2009;131:5214-23.
[62] Ganguly D, Chen J. Modulation of the Disordered Conformational Ensembles of the p53 Transactivation Domain by Cancer-Associated Mutations. 2015.
[63] Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc Natl Acad Sci U S A. 2010;107:8183-8.
[64] Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci U S A. 2013;110:13392-7.
[65] Chen JH. Intrinsically Disordered p53 Extreme C-Terminus Binds to S100B(beta beta) through "Fly-Casting". Journal of the American Chemical Society. 2009;131:2088-+.
[66] Karanicolas J, Brooks CL, 3rd. The origins of asymmetry in the folding transition states of protein L and protein G. Protein science : a publication of the Protein Society. 2002;11:2351-61.
[67] Frauenfelder H, Sligar SG, Wolynes PG. The Energy Landscapes and Motions of Proteins. Science. 1991;254:1598-603.
[68] Wolynes PG, Onuchic JN, Thirumalai D. Navigating the Folding Routes. Science. 1995;267:1619-20.
[69] Ganguly D, Otieno S, Waddell B, Iconaru L, Kriwacki RW, Chen J. Electrostatically accelerated coupled binding and folding of intrinsically disordered proteins. J Mol Biol. 2012;422:674-84.
[70] Turjanski AG, Gutkind JS, Best RB, Hummer G. Binding-induced folding of a natively unstructured transcription factor. PLoS Comput Biol. 2008;4:e1000060.

[71] Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. Proc Natl Acad Sci U S A. 2018;115:E4758-E66.
[72] Lee KH, Chen JH. Optimization of the GBMV2 implicit solvent force field for accurate simulation of protein conformational equilibria. Journal of computational chemistry. 2017;38:1332-41.
[73] Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nature methods. 2017;14:71-3.
[74] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. 2010;78:1950-8.
[75] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins. 2006;65:712-25.
[76] Piana S, Donchev AG, Robustelli P, Shaw DE. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. J Phys Chem B. 2015;119:5113-23.
[77] Best RB, Zheng W, Mittal J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. J Chem Theory Comput. 2014;10:5113-24.
[78] Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett. 1999;314:141-51.
[79] Earl DJ, Deem MW. Parallel tempering: Theory, applications, and new perspectives. Phys Chem Chem Phys. 2005;7:3910-6.
[80] McDowell C, Chen J, Chen J. Potential conformational heterogeneity of p53 bound to S100B (ββ). J Mol Biol. 2013;425:999-1010.
[81] Zhang W, Ganguly D, Chen J. Residual structures, conformational fluctuations, and electrostatic interactions in the synergistic folding of two intrinsically disordered proteins. Plos Computational Biology. 2012;8:e1002353.
[82] Rao F, Caflisch A. Replica exchange molecular dynamics simulations of reversible folding. J Chem Phys. 2003;119:4035-42.
[83] Zhang WH, Chen JH. Accelerate Sampling in Atomistic Energy Landscapes Using Topology-Based Coarse-Grained Models. J Chem Theory Comput. 2014;10:918-23.
[84] Lee MS, Olson MA. Comparison of two adaptive temperature-based replica exchange methods applied to a sharp phase transition of protein unfolding-folding. J Chem Phys. 2011;134.
[85] Trebst S, Troyer M, Hansmann UHE. Optimized parallel tempering simulations of proteins. J Chem Phys. 2006;124.
[86] Spiriti J, Kamberaj H, Van Der Vaart A. Development and Application of Enhanced Sampling Techniques to Simulate the Long-Time Scale Dynamics of Biomolecular Systems. Int J Quantum Chem. 2012;112:33-43.
[87] Kamberaj H, van der Vaart A. An optimized replica exchange molecular dynamics method. J Chem Phys. 2009;130.
[88] Spill YG, Bouvier G, Nilges M. A convective replica-exchange method for sampling new energy basins. Journal of computational chemistry. 2013;34:132-40.

[89] Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. J Chem Phys. 2004;120:11919-29.
[90] Laio A, Parrinello M. Escaping free-energy minima. Proc Natl Acad Sci U S A. 2002;99:12562-6.
[91] Valsson O, Parrinello M. Variational Approach to Enhanced Sampling and Free Energy Calculations. Physical review letters. 2014;113.
[92] Zheng LQ, Chen MG, Yang W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. Proc Natl Acad Sci U S A. 2008;105:20227-32.
[93] Bicout DJ, Szabo A. Entropic barriers, transition states, funnels, and exponential protein folding kinetics: A simple model. Protein Science. 2000;9:452-65.
[94] Liu XR, Beugelsdijk A, Chen JH. Dynamics of the BH3-Only Protein Binding Interface of Bcl-xL. Biophys J. 2015;109:1049-57.
[95] Liu XR, Jia ZG, Chen JH. Enhanced Sampling of Intrinsic Structural Heterogeneity of the BH3-Only Protein Binding Interface of Bcl-xL. J Phys Chem B. 2017;121:9160-8.
[96] Liu XR, Chen JL, Chen JH. Residual Structure Accelerates Binding of Intrinsically Disordered ACTR by Promoting Efficient Folding upon Encounter. J Mol Biol. 2019;431:422-32.
[97] Liu XR, Chen JH. HyRes: a coarse-grained model for multi-scale enhanced sampling of disordered protein conformations. Phys Chem Chem Phys. 2017;19:32421-32.
[98] Tsai CJ, Ma B, Sham YY, Kumar S, Nussinov R. Structured disorder and conformational selection. Proteins. 2001;44:418-27.
[99] Zhang C, Ma J. Folding helical proteins in explicit solvent using dihedral-biased tempering. Proc Natl Acad Sci U S A. 2012;109:8139-44.
[100] Wright PE, Dyson HJ. Linking folding and binding. Curr Opin Struc Biol. 2009;19:31-8.
[101] Lyle N, Das RK, Pappu RV. A quantitative measure for protein conformational heterogeneity. J Chem Phys. 2013;139:121907.
[102] Zhou HX. Intrinsic disorder: signaling via highly specific but short-lived association. Trends Biochem Sci. 2012;37:43-8.
[103] Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem Sci. 2008;33:2-8.
[104] McDowell C, Chen J, Chen J. Potential Conformational Heterogeneity of p53 Bound to S100B(betabeta). J Mol Biol. 2013.
[105] Mittag T, Orlicky S, Choy WY, Tang XJ, Lin H, Sicheri F, et al. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. Proc Natl Acad Sci U S A. 2008;105:17772-7.
[106] Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, et al. Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase. Structure. 2010;18:494-506.
[107] Mitrea DM, Kriwacki RW. Regulated unfolding of proteins in signaling. Febs Lett. 2013;587:1081-8.

[108] Grimmler M, Wang YF, Mund T, Cilensek Z, Keidel EM, Waddell MB, et al. Cdk-inhibitory activity and stability of p27(Kip1) are directly regulated by oncogenic tyrosine kinases. Cell. 2007;128:269-80.
[109] Follis AV, Chipuk JE, Fisher JC, Yun MK, Grace CR, Nourse A, et al. PUMA binding induces partial unfolding within BCL-xL to disrupt p53 binding and promote apoptosis. Nat Chem Biol. 2013;9:163-8.
[110] Denisov AY, Madiraju MS, Chen G, Khadir A, Beauparlant P, Attardo G, et al. Solution structure of human BCL-w: modulation of ligand binding by the C-terminal helix. J Biol Chem. 2003;278:21124-8.
[111] Hinds MG, Lackmann M, Skea GL, Harrison PJ, Huang DC, Day CL. The structure of Bcl-w reveals a role for the C-terminal residues in modulating biological activity. EMBO J. 2003;22:1497-507.
[112] Bardwell JCA, Jakob U. Conditional disorder in chaperone action. Trends Biochem Sci. 2012;37:517-25.
[113] Foit L, George JS, Zhang BW, Brooks CL, 3rd, Bardwell JC. Chaperone activation by unfolding. Proc Natl Acad Sci U S A. 2013;110:E1254-62.
[114] Chen J, Zolkiewska A. Force-induced unfolding simulations of the human Notch1 negative regulatory region: possible roles of the heterodimerization domain in mechanosensing. PLoS One. 2011;6:e22837.
[115] Stephenson NL, Avis JM. Direct observation of proteolytic cleavage at the S2 site upon forced unfolding of the Notch negative regulatory region. Proc Natl Acad Sci U S A. 2012;109:E2757-65.
[116] Yip KW, Reed JC. Bcl-2 family proteins and cancer. Oncogene. 2008;27:6398-406.
[117] Petros AM, Olejniczak ET, Fesik SW. Structural biology of the Bcl-2 family of proteins. Bba-Mol Cell Res. 2004;1644:83-94.
[118] Green DR, Kroemer G. Cytoplasmic functions of the tumour suppressor p53. Nature. 2009;458:1127-30.
[119] Follis AV, Llambi F, Ou L, Baran K, Green DR, Kriwacki RW. The DNA-binding domain mediates both nuclear and cytosolic functions of p53. Nat Struct Mol Biol. 2014;21:535-43.
[120] Michels J, Kepp O, Senovilla L, Lissa D, Castedo M, Kroemer G, et al. Functions of BCL-X L at the interface between cell death and metabolism. International journal of cell biology. 2013;2013.
[121] Chipuk JE, Moldoveanu T, Llambi F, Parsons MJ, Green DR. The BCL-2 family reunion. Mol Cell. 2010;37:299-310.
[122] Chipuk JE, Green DR. How do BCL-2 proteins induce mitochondrial outer membrane permeabilization? Trends in cell biology. 2008;18:157-64.
[123] Hinds MG, Smits C, Fredericks-Short R, Risk JM, Bailey M, Huang DC, et al. Bim, Bad and Bmf: intrinsically unstructured BH3-only proteins that undergo a localized conformational change upon binding to prosurvival Bcl-2 targets. Cell Death Differ. 2007;14:128-36.
[124] Chipuk JE, Bouchier-Hayes L, Kuwana T, Newmeyer DD, Green DR. PUMA couples the nuclear and cytoplasmic proapoptotic function of p53. Science. 2005;309:1732-5.

[125] Oltersdorf T, Elmore SW, Shoemaker AR, Armstrong RC, Augeri DJ, Belli BA, et al. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. Nature. 2005;435:677-81.
[126] Priyadarshi A, Roy A, Kim K-S, Kim EE, Hwang KY. Structural insights into mouse anti-apoptotic Bcl-xl reveal affinity for Beclin 1 and gossypol. Biochem Bioph Res Co. 2010;394:515-21.
[127] Manion MK, O'Neill JW, Giedt CD, Kim KM, Zhang KY, Hockenbery DM. Bcl-XL mutations suppress cellular sensitivity to antimycin A. J Biol Chem. 2004;279:2159-65.
[128] Aritomi M, Kunishima N, Inohara N, Ishibashi Y, Ohta S, Morikawa K. Crystal structure of rat Bcl-XL Implications for the function of the Bcl-2 protein family. J Biol Chem. 1997;272:27886-92.
[129] X L, S D, Y Z, P M, JW K. The structure of a Bcl-xL/Bim fragment complex: implications for Bim function. Immunity. 2003;19:341-52.
[130] Feng Y, Zhang L, Hu T, Shen X, Ding J, Chen K, et al. A conserved hydrophobic core at Bcl-x_L mediates its structural stability and binding affinity with BH3-domain peptide of pro-apoptotic protein. Archives of biochemistry and biophysics. 2009;484:46-54.
[131] Muchmore SW, Sattler M, Liang H, Meadows RP, Harlan JE, Yoon HS, et al. X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. Nature. 1996;381:335-41.
[132] Lessene G, Czabotar PE, Sleebs BE, Zobel K, Lowes KN, Adams JM, et al. Structure-guided design of a selective BCL-XL inhibitor. Nature chemical biology. 2013;9:390-7.
[133] Feng W, Huang S, Wu H, Zhang M. Molecular basis of Bcl-xL's target recognition versatility revealed by the structure of Bcl-xL in complex with the BH3 domain of Beclin-1. J Mol Biol. 2007;372:223-35.
[134] Sattler M, Liang H, Nettesheim D, Meadows RP, Harlan JE, Eberstadt M, et al. Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. Science. 1997;275:983-6.
[135] Boersma MD, Haase HS, Peterson-Kaufman KJ, Lee EF, Clarke OB, Colman PM, et al. Evaluation of diverse α/β-backbone patterns for functional α-helix mimicry: analogues of the Bim BH3 domain. Journal of the American Chemical Society. 2011;134:315-23.
[136] Zhou H, Chen J, Meagher JL, Yang C-Y, Aguilar A, Liu L, et al. Design of Bcl-2 and Bcl-xL inhibitors with subnanomolar binding affinities based upon a new scaffold. Journal of medicinal chemistry. 2012;55:4664-82.
[137] Lee E, Czabotar P, Smith B, Deshayes K, Zobel K, Colman P, et al. Crystal structure of ABT-737 complexed with Bcl-xL: implications for selectivity of antagonists of the Bcl-2 family. 2007.
[138] Bruncko M, Oost TK, Belli BA, Ding H, Joseph MK, Kunzer A, et al. Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-xL. Journal of medicinal chemistry. 2007;50:641-62.
[139] Schroeder GM, Wei D, Banfi P, Cai Z-W, Lippy J, Menichincheri M, et al. Pyrazole and pyrimidine phenylacylsulfonamides as dual Bcl-2/Bcl-xL antagonists. Bioorganic & medicinal chemistry letters. 2012;22:3951-6.

[140] Sleebs BE, Czabotar PE, Fairbrother WJ, Fairlie WD, Flygare JA, Huang DC, et al. Quinazoline sulfonamides as dual binders of the proteins B-cell lymphoma 2 and B-cell lymphoma extra long with potent proapoptotic cell-based activity. Journal of medicinal chemistry. 2011;54:1914-26.
[141] Wysoczanski P, Mart RJ, Loveridge EJ, Williams C, Whittaker SB-M, Crump MP, et al. NMR solution structure of a photoswitchable apoptosis activating Bak peptide bound to Bcl-xL. Journal of the American Chemical Society. 2012;134:7644-7.
[142] Okamoto T, Zobel K, Fedorova A, Quan C, Yang H, Fairbrother WJ, et al. Stabilizing the pro-apoptotic BimBH3 helix (BimSAHB) does not necessarily enhance affinity or biological activity. ACS chemical biology. 2012;8:297-302.
[143] Lee EF, Sadowsky JD, Smith BJ, Czabotar PE, Peterson-Kaufman KJ, Colman PM, et al. High-Resolution Structural Characterization of a Helical α/β-Peptide Foldamer Bound to the Anti-Apoptotic Protein Bcl-xL. Angewandte Chemie International Edition. 2009;48:4318-22.
[144] Petros AM, Nettesheim DG, Wang Y, Olejniczak ET, Meadows RP, Mack J, et al. Rationale for Bcl‐XL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. Protein Science. 2000;9:2528-34.
[145] Czabotar PE, Lee EF, Thompson GV, Wardak AZ, Fairlie WD, Colman PM. Mutation to Bax beyond the BH3 domain disrupts interactions with pro-survival proteins and promotes apoptosis. J Biol Chem. 2011;286:7123-31.
[146] Feig M, Karanicolas J, Brooks CL. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Graph Model. 2004;22:377-95.
[147] Petros AM, Nettesheim DG, Wang Y, Olejniczak ET, Meadows RP, Mack J, et al. Rationale for Bcl-xL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. Protein Sci. 2000;9:2528-34.
[148] Lee EF, Sadowsky JD, Smith BJ, Czabotar PE, Peterson-Kaufman KJ, Colman PM, et al. High-resolution structural characterization of a helical alpha/beta-peptide foldamer bound to the anti-apoptotic protein Bcl-xL. Angew Chem Int Ed Engl. 2009;48:4318-22.
[149] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. Journal of computational chemistry. 1983;4:187-217.
[150] Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: The Biomolecular Simulation Program. Journal of computational chemistry. 2009;30:1545-614.
[151] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem. 2005;26:1781-802.
[152] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998;102:3586-616.
[153] Mackerell AD, Feig M, Brooks CL. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. Journal of computational chemistry. 2004;25:1400-15.
[154] Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, et al. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the

Backbone ϕ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. J Chem Theory Comput. 2012;8:3257-73.
[155] Darden T, York D, Pedersen L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. J Chem Phys. 1993;98:10089-92.
[156] Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. J Comput Phys. 1977;23:327-41.
[157] Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph. 1996;14:33-&.
[158] Berjanskii M, Wishart DS. NMR: prediction of protein flexibility. Nat Protoc. 2006;1:683-8.
[159] Ishida T, Kinoshita K. Prediction of disordered regions in proteins based on the meta approach. Bioinformatics. 2008;24:1344-8.
[160] Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22:2577-637.
[161] Rogers JM, Oleinikovas V, Shammas SL, Wong CT, De Sancho D, Baker CM, et al. Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein. Proc Natl Acad Sci U S A. 2014;111:15420-5.
[162] Lane TJ, Shukla D, Beauchamp KA, Pande VS. To milliseconds and beyond: challenges in the simulation of protein folding. Curr Opin Struct Biol. 2013;23:58-65.
[163] Zhang W, Chen J. Accelerate Sampling in Atomistic Energy Landscapes Using Topology-Based Coarse-Grained Models. J Chem Theory Comput. 2014;10:918-23.
[164] Zuckerman DM. Equilibrium Sampling in Biomolecular Simulations. In: Rees DC, Dill KA, Williamson JR, editors. Annual Review of Biophysics, Vol 402011. p. 41-62.
[165] Jeng PS, Cheng EH. Cancer therapeutics: Pulling the plug on BCL-X(L). Nat Chem Biol. 2013;9:351-2.
[166] Chen J. Towards the physical basis of how intrinsic disorder mediates protein function. Arch Biochem Biophys. 2012;524:123-31.
[167] Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. Intrinsic protein disorder in complete genomes. Genome informatics Workshop on Genome Informatics. 2000;11:161-71.
[168] Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. Journal of biomolecular structure & dynamics. 2012;30:137-49.
[169] Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry. 2002;41:6573-82.
[170] Babu MM, van der Lee R, de Groot NS, Gsponer J. Intrinsically disordered proteins: regulation and disease. Curr Opin Struct Biol. 2011;21:432-40.
[171] Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition (vol 5, pg 789, 2009). Nature chemical biology. 2009;5:954-.
[172] Chiti F, Dobson CM. Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem. 2006;75:333-66.
[173] Lu HC, Chung SS, Fornili A, Fraternali F. Anatomy of protein disorder, flexibility and disease-related mutations. Frontiers in molecular biosciences. 2015;2:47.

[174] Vacic V, Markwick PR, Oldfield CJ, Zhao X, Haynes C, Uversky VN, et al. Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. PLoS Comput Biol. 2012;8:e1002709.
[175] Gsponer J, Futschik ME, Teichmann SA, Babu MM. Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation. Science. 2008;322:1365-8.
[176] Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. Cell. 2009;138:198-208.
[177] Uversky VN. Intrinsic Disorder-based Protein Interactions and their Modulators. Curr Pharm Design. 2013;19:4191-213.
[178] Teilum K, Olsen JG, Kragelund BB. Functional aspects of protein flexibility. Cellular and molecular life sciences : CMLS. 2009;66:2231-47.
[179] Shammas SL, Crabtree MD, Dahal L, Wicky BIM, Clarke J. Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies. J Biol Chem. 2016;291:6689-95.
[180] Czabotar PE, Lessene G, Strasser A, Adams JM. Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. Nat Rev Mol Cell Bio. 2014;15:49-63.
[181] Follis AV, Chipuk JE, Fisher JC, Yun M-K, Grace CR, Nourse A, et al. PUMA binding induces partial unfolding within BCL-xL to disrupt p53 binding and promote apoptosis. Nat Chem Biol. 2013;9:163-8.
[182] Hinds M, Smits C, Fredericks-Short R, Risk J, Bailey M, Huang D, et al. Bim, Bad and Bmf: intrinsically unstructured BH3-only proteins that undergo a localized conformational change upon binding to prosurvival Bcl-2 targets. Cell Death Differ. 2006;14:128-36.
[183] Liu P, Kim B, Friesner RA, Berne BJ. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. Proc Natl Acad Sci U S A. 2005;102:13749-54.
[184] Wang LL, Friesner RA, Berne BJ. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). J Phys Chem B. 2011;115:9431-8.
[185] Hansmann UHE. Parallel tempering algorithm for conformational studies of biological molecules. Chem Phys Lett. 1997;281:140-50.
[186] Pang X, Zhou HX. Disorder-to-Order Transition of an Active-Site Loop Mediates the Allosteric Activation of Sortase A. Biophys J. 2015;109:1706-15.
[187] Huang K, Garcia AE. Acceleration of Lateral Equilibration in Mixed Lipid Bilayers Using Replica Exchange with Solute Tempering. J Chem Theory Comput. 2014;10:4264-72.
[188] Terakawa T, Kameda T, Takada S. On Easy Implementation of a Variant of the Replica Exchange with Solute Tempering in GROMACS. Journal of computational chemistry. 2011;32:1228-34.
[189] Stirnemann G, Sterpone F. Recovering Protein Thermal Stability Using All-Atom Hamiltonian Replica-Exchange Simulations in Explicit Solvent. J Chem Theory Comput. 2015;11:5573-7.

[190] Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. 2013;29:845-54.
[191] Bussi G. Hamiltonian replica exchange in GROMACS: a flexible implementation. Mol Phys. 2014;112:379-84.
[192] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. Journal of computational chemistry. 1997;18:1463-72.
[193] Fletcher CM, Jones DNM, Diamond R, Neuhaus D. Treatment of NOE constraints involving equivalent or nonstereoassigned protons in calculations of biomacromolecular structures. Journal of biomolecular NMR. 1996;8:292-310.
[194] McDowell C, Chen JL, Chen JH. Potential Conformational Heterogeneity of p53 Bound to S100B(beta beta). J Mol Biol. 2013;425:999-1010.
[195] Petros AM, Nettesheim DG, Wang Y, Olejniczak ET, Meadows RP, Mack J, et al. Rationale for Bcl-xL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. Protein science : a publication of the Protein Society. 2000;9:2528-34.
[196] Bohnuud T, Kozakov D, Vajda S. Evidence of conformational selection driving the formation of ligand binding sites in protein-protein interfaces. PLoS Comput Biol. 2014;10:e1003872.
[197] Knott M, Best RB. A preformed binding interface in the unbound ensemble of an intrinsically disordered protein: evidence from molecular simulations. Plos Computational Biology. 2012;8:e1002605.
[198] Ganguly D, Zhang W, Chen J. Synergistic folding of two intrinsically disordered proteins: searching for conformational selection. Molecular BioSystems. 2012;8:198-209.
[199] Click TH, Ganguly D, Chen J. Intrinsically Disordered Proteins in a Physics-Based World. Int J Mol Sci. 2010;11:5292-309.
[200] Das RK, Ruff KM, Pappu RV. Relating sequence encoded information to form and function of intrinsically disordered proteins. Curr Opin Struct Biol. 2015;32:102-12.
[201] Arai M, Sugase K, Dyson HJ, Wright PE. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. Proc Natl Acad Sci U S A. 2015;112:9614-9.
[202] Toto A, Camilloni C, Giri R, Brunori M, Vendruscolo M, Gianni S. Molecular Recognition by Templated Folding of an Intrinsically Disordered Protein. Sci Rep. 2016;6:21994.
[203] Rogers JM, Wong CT, Clarke J. Coupled Folding and Binding of the Disordered Protein PUMA Does Not Require Particular Residual Structure. Journal of the American Chemical Society. 2014;136:5197-200.
[204] Crabtree MD, Mendonca C, Bubb QR, Clarke J. Folding and binding pathways of BH3-only proteins are encoded within their intrinsically disordered sequence, not templated by partner proteins. J Biol Chem. 2018;293:9718-23.
[205] Dahal L, Kwan TOC, Shammas SL, Clarke J. pKID Binds to KIX via an Unstructured Transition State with Nonnative Interactions. Biophys J. 2017;113:2713-22.
[206] Fisher CK, Stultz CM. Constructing ensembles for intrinsically disordered proteins. Curr Opin Struc Biol. 2011;21:426-31.

[207] Borcherds W, Theillet FX, Katzer A, Finzel A, Mishall KM, Powell AT, et al. Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. Nature chemical biology. 2014;10:1000-2.
[208] Bienkiewicz EA, Adkins JN, Lumb KJ. Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27(Kip1). Biochemistry. 2002;41:752-9.
[209] Ganguly D, Otieno S, Waddell B, Iconaru L, Kriwacki RW, Chen J. Electrostatically Accelerated Coupled Binding and Folding of Intrinsically Disordered Proteins. J Mol Biol. 2012;422:674-84.
[210] Ganguly D, Zhang W, Chen J. Electrostatically Accelerated Encounter and Folding for Facile Recognition of Intrinsically Disordered Proteins. PLoS Comput Biol. 2013;9:e1003363.
[211] Zhou HX. From Induced Fit to Conformational Selection: A Continuum of Binding Mechanism Controlled by the Timescale of Conformational Transitions. Biophys J. 2010;98:L15-L7.
[212] Kubelka J, Hofrichter J, Eaton WA. The protein folding 'speed limit'. Curr Opin Struct Biol. 2004;14:76-88.
[213] Huang Y, Liu Z. Kinetic advantage of intrinsically disordered proteins in coupled folding–binding process: a critical assessment of the “fly-casting” mechanism. J Mol Biol. 2009;393:1143-59.
[214] Dogan J, Gianni S, Jemth P. The binding mechanisms of intrinsically disordered proteins. Phys Chem Chem Phys. 2014;16:6323-31.
[215] Arai M, Sugase K, Dyson HJ, Wright PE. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. Proceedings of the National Academy of sciences. 2015;112:9614-9.
[216] Giri R, Morrone A, Toto A, Brunori M, Gianni S. Structure of the transition state for the binding of c-Myb and KIX highlights an unexpected order for a disordered system. Proc Natl Acad Sci U S A. 2013;110:14942-7.
[217] Hill SA, Kwa LG, Shammas SL, Lee JC, Clarke J. Mechanism of Assembly of the Non-Covalent Spectrin Tetramerization Domain from Intrinsically Disordered Partners. J Mol Biol. 2014;426:21-35.
[218] Demarest SJ, Martinez-Yamout M, Chung J, Chen HW, Xu W, Dyson HJ, et al. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. Nature. 2002;415:549-53.
[219] Ganguly D, Chen J. Topology-based modeling of intrinsically disordered proteins: balancing intrinsic folding and intermolecular interactions. Proteins. 2011;79:1251-66.
[220] Huang Y, Liu Z. Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. J Mol Biol. 2009;393:1143-59.
[221] Wolynes PG. Recent successes of the energy landscape theory of protein folding and function. Q Rev Biophys. 2005;38:405-10.
[222] Lu Q, Lu HP, Wang J. Exploring the mechanism of flexible biomolecular recognition with single molecule dynamics. Phys Rev Lett. 2007;98:128105.
[223] Wang J, Wang Y, Chu X, Hagen SJ, Han W, Wang E. Multi-Scaled Explorations of Binding-Induced Folding of Intrinsically Disordered Protein Inhibitor IA3 to its Target Enzyme. PLoS Comput Biol. 2011;7:e1001118.

[224] Law SM, Gagnon JK, Mapp AK, Brooks CL, 3rd. Prepaying the entropic cost for allosteric regulation in KIX. Proc Natl Acad Sci U S A. 2014;111:12067-72.
[225] Law SM, Ahlstrom LS, Panahi A, Brooks III CL. Hamiltonian Mapping Revisited: Calibrating Minimalist Models to Capture Molecular Recognition by Intrinsically Disordered Proteins. The journal of physical chemistry letters. 2014;5:3441-4.
[226] Shea JE, Nochomovitz YD, Guo ZY, Brooks CL. Exploring the space of protein folding Hamiltonians: The balance of forces in a minimalist beta-barrel model. J Chem Phys. 1998;109:2895-903.
[227] Wolynes PG. Evolution, energy landscapes and the paradoxes of protein folding. Biochimie. 2015;119:218-30.
[228] Zhang BW, Jasnow D, Zuckerman DM. Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin. Proceedings of the National Academy of sciences. 2007;104:18043-8.
[229] Chung HS, Louis JM, Eaton WA. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. Proc Natl Acad Sci U S A. 2009;106:11837-44.
[230] Neupane K, Ritchie DB, Yu H, Foster DAN, Wang F, Woodside MT. Transition Path Times for Nucleic Acid Folding Determined from Energy-Landscape Analysis of Single-Molecule Trajectories. Physical review letters. 2012;109.
[231] Dogan J, Mu X, Engstrom A, Jemth P. The transition state structure for coupled binding and folding of disordered protein domains. Sci Rep-Uk. 2013;3.
[232] Chu WT, Clarke J, Shammas SL, Wang J. Role of non-native electrostatic interactions in the coupled folding and binding of PUMA with Mcl-1. PLoS computational biology. 2017;13.
[233] Bryngelson JD, Wolynes PG. Spin glasses and the statistical mechanics of protein folding. Proceedings of the National Academy of sciences. 1987;84:7524-8.
[234] Onuchic JN, Wolynes PG. Theory of protein folding. Curr Opin Struc Biol. 2004;14:70-5.
[235] Karanicolas J, Brooks CL. The origins of asymmetry in the folding transition states of protein L and protein G. Protein Sci. 2002;11:2351-61.
[236] Dogan J, Schmidt T, Mu X, Engstrom A, Jemth P. Fast association and slow transitions in the interaction between two intrinsically disordered protein domains. J Biol Chem. 2012;287:34316-24.
[237] Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci. 2010;35:539-46.
[238] Karanicolas J, Brooks CL. Improved Go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. J Mol Biol. 2003;334:309-25.
[239] Law SM, Ahlstrom LS, Panahi A, Brooks CL, 3rd. Hamiltonian Mapping Revisited: Calibrating Minimalist Models to Capture Molecular Recognition by Intrinsically Disordered Proteins. The journal of physical chemistry letters. 2014;5:3441-4.
[240] Uversky VN, Dave V, Iakoucheva LM, Malaney P, Metallo SJ, Pathak RR, et al. Pathological Unfoldomics of Uncontrolled Chaos: Intrinsically Disordered Proteins and Human Diseases. Chemical reviews. 2014;114:6844-79.

[241] Kruschel D, Zagrovic B. Conformational averaging in structural biology: issues, challenges and computational solutions. Molecular BioSystems. 2009;5:1606-16.
[242] Moritsugu K, Terada T, Kidera A. Scalable free energy calculation of proteins via multiscale essential sampling. J Chem Phys. 2010;133:224105.
[243] Lee KH, Chen J. Multiscale enhanced sampling of intrinsically disordered protein conformations. Journal of computational chemistry. 2015.
[244] Lwin TZ, Luo R. Overcoming entropic barrier with coupled sampling at dual resolutions. The Journal of chemical physics. 2005;123:194904.
[245] Liu P, Shi Q, Lyman E, Voth GA. Reconstructing atomistic detail for coarse-grained models with resolution exchange. J Chem Phys. 2008;129:114103.
[246] Lyman E, Zuckerman DM. Resolution Exchange Simulation with Incremental Coarsening. J Chem Theory Comput. 2006;2:656-66.
[247] Liwo A, Khalili M, Czaplewski C, Kalinowski S, Oldziej S, Wachucik K, et al. Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins. J Phys Chem B. 2007;111:260-85.
[248] Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. A united-residue force field for off-lattice protein-structure simulations .1. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. Journal of computational chemistry. 1997;18:849-73.
[249] Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Oldziej S, Scheraga HA. A united-residue force field for off-lattice protein-structure simulations .2. Parameterization of short-range interactions and determination of weights of energy terms by Z-score optimization. Journal of computational chemistry. 1997;18:874-87.
[250] Oldziej S, Czaplewski C, Liwo A, Chinchio M, Nanias M, Vila JA, et al. Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: Assessment in two blind tests. Proc Natl Acad Sci U S A. 2005;102:7547-52.
[251] Liwo A, Khalili M, Scheraga HA. Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. Proc Natl Acad Sci U S A. 2005;102:2362-7.
[252] Maisuradze GG, Senet P, Czaplewski C, Liwo A, Scheraga HA. Investigation of Protein Folding by Coarse-Grained Molecular Dynamics with the UNRES Force Field. J Phys Chem A. 2010;114:4471-85.
[253] Smith AV, Hall CK. alpha-helix formation: Discontinuous molecular dynamics on an intermediate-resolution protein model. Proteins-Structure Function and Genetics. 2001;44:344-60.
[254] Cheon M, Chang I, Hall CK. Extending the PRIME model for protein aggregation to all 20 amino acids. Proteins. 2010;78:2950-60.
[255] Nguyen HD, Hall CK. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. Proc Natl Acad Sci U S A. 2004;101:16180-5.
[256] Nguyen HD, Hall CK. Spontaneous fibril formation by polyalanines; Discontinuous molecular dynamics simulations. Journal of the American Chemical Society. 2006;128:1890-901.
[257] Bereau T, Deserno M. Generic coarse-grained model for protein folding and aggregation. J Chem Phys. 2009;130.

[258] Maupetit J, Tuffery P, Derreumaux P. A coarse-grained protein force field for folding and structure prediction. Proteins. 2007;69:394-408.
[259] Derreumaux P. From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential. J Chem Phys. 1999;111:2301-10.
[260] Chebaro Y, Derreumaux P. Targeting the early steps of A β 16 – 22 protofibril disassembly by N - methylated inhibitors: A numerical study. Proteins: Structure, Function, and Bioinformatics. 2009;75:442-52.
[261] Chebaro Y, Pasquali S, Derreumaux P. The Coarse-Grained OPEP Force Field for Non-Amyloid and Amyloid Proteins. J Phys Chem B. 2012;116:8741-52.
[262] Kynast P, Derreumaux P, Strodel B. Evaluation of the coarse-grained OPEP force field for protein-protein docking. Bmc Biophys. 2016;9.
[263] Kar P, Gopal SM, Cheng YM, Predeus A, Feig M. PRIMO: A Transferable Coarse-Grained Force Field for Proteins. J Chem Theory Comput. 2013;9:3769-88.
[264] Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI coarse-grained force field: Extension to proteins. J Chem Theory Comput. 2008;4:819-34.
[265] de Jong DH, Singh G, Bennett WFD, Arnarez C, Wassenaar TA, Schafer LV, et al. Improved Parameters for the Martini Coarse-Grained Protein Force Field. J Chem Theory Comput. 2013;9:687-97.
[266] Han W, Schulten K. Further Optimization of a Hybrid United-Atom and Coarse-Grained Force Field for Folding Simulations: Improved Backbone Hydration and Interactions between Charged Side Chains. J Chem Theory Comput. 2012;8:4413-24.
[267] Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. Coarse-Grained Protein Models and Their Applications. Chemical reviews. 2016;116:7898-936.
[268] Ingolfsson HI, Lopez CA, Uusitalo JJ, de Jong DH, Gopal SM, Periole X, et al. The power of coarse graining in biomolecular simulations. Wiley interdisciplinary reviews Computational molecular science. 2014;4:225-48.
[269] Marrink SJ, Tieleman DP. Perspective on the Martini model. Chemical Society reviews. 2013;42:6801-22.
[270] Sterpone F, Melchionna S, Tuffery P, Pasquali S, Mousseau N, Cragnolini T, et al. The OPEP protein model: from single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA systems. Chemical Society reviews. 2014;43:4871-93.
[271] Jia ZG, Chen JH. Necessity of High-Resolution for Coarse-Grained Modeling of Flexible Proteins. Journal of computational chemistry. 2016;37:1725-33.
[272] Parigi G, Rezaei-Ghaleh N, Giachetti A, Becker S, Fernandez C, Blackledge M, et al. Long-range correlated dynamics in intrinsically disordered proteins. J Am Chem Soc. 2014;136:16201-9.
[273] Jia Z, Chen J. Necessity of high-resolution for coarse-grained modeling of flexible proteins. J Comput Chem. 2016;37:1725-33.
[274] MacKerell AD, Feig M, Brooks CL. Improved treatment of the protein backbone in empirical force fields. Journal of the American Chemical Society. 2004;126:698-9.
[275] Neria E, Fischer S, Karplus M. Simulation of activation free energies in molecular systems. J Chem Phys. 1996;105:1902-21.
[276] Pierson NA, Chen LX, Russell DH, Clemmer DE. Cis-Trans Isomerizations of Proline Residues Are Key to Bradykinin Conformations. Journal of the American Chemical Society. 2013;135:3186-92.

[277] Miyazawa S, Jernigan RL. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol. 1996;256:623-44.
[278] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1:19-25.
[279] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. J Chem Phys. 1983;79:926-35.
[280] Darden T, York D, Pedersen L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. The Journal of chemical physics. 1993;98:10089-92.
[281] Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. Journal of computational chemistry. 1992;13:1011-21.
[282] Fesinmeyer RM, Hudson FM, Andersen NH. Enhanced hairpin stability through loop design: The case of the protein G B1 domain hairpin. Journal of the American Chemical Society. 2004;126:7238-43.
[283] Radhakrishnan I, PerezAlvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:Coactivator interactions. Cell. 1997;91:741-52.
[284] Demarest SJ, Martinez-Yamout M, Chung J, Chen H, Xu W, Dyson HJ, et al. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. Nature. 2002;415:549-53.
[285] Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci U S A. 2013;110:13392-7.
[286] Clementi C, Garcia AE, Onuchic JN. Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: All-atom representation study of protein L. J Mol Biol. 2003;326:933-54.
[287] Chakrabartty A, Kortemme T, Baldwin RL. Helix Propensities of the Amino-Acids Measured in Alanine-Based Peptides without Helix-Stabilizing Side-Chain Interactions. Protein Science. 1994;3:843-52.
[288] Pace CN, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. Biophys J. 1998;75:422-7.
[289] Drake JA, Pettitt BM. Force field-dependent solution properties of glycine oligomers. J Comput Chem. 2015;36:1275-85.
[290] Chen JH. Effective Approximation of Molecular Volume Using Atom-Centered Dielectric Functions in Generalized Born Models. J Chem Theory Comput. 2010;6:2790-803.
[291] Kim YC, Hummer G. Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. J Mol Biol. 2008;375:1416-33.
[292] Shalongo W, Dugad L, Stellwagen E. Distribution of helicity within the model peptide acetyl (AAQAA) 3amide. Journal of the American Chemical Society. 1994;116:8288-93.
[293] Best RB. Atomistic molecular simulations of protein folding. Curr Opin Struc Biol. 2012;22:52-61.

[294] Radhakrishnan I, Perez-Alvarado GC, Dyson HJ, Wright PE. Conformational preferences in the Ser(133)-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. FEBS Lett. 1998;430:317-22.
[295] Xiang S, Gapsys V, Kim HY, Bessonov S, Hsiao HH, Mohlmann S, et al. Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. Structure. 2013;21:2162-74.
[296] Rauscher S, Gapsys V, Gajda MJ, Zweckstetter M, de Groot BL, Grubmuller H. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. J Chem Theory Comput. 2015;11:5513-24.
[297] Demarest SJ, Deechongkit S, Dyson HJ, Evans RM, Wright PE. Packing, specificity, and mutability at the binding interface between the p160 coactivator and CREB-binding protein. Protein Sci. 2004;13:203-10.
[298] Ebert MO, Bae SH, Dyson HJ, Wright PE. NMR relaxation study of the complex formed between CBP and the activation domain of the nuclear hormone receptor coactivator ACTR. Biochemistry. 2008;47:1299-308.
[299] Kjaergaard M, Nørholm A-B, Hendus-Altenburger R, Pedersen SF, Poulsen FM, Kragelund BB. Temperature-dependent structural changes in intrinsically disordered proteins: Formation of α-helices or loss of polyproline II? Protein Science. 2010;19:1555-64.
[300] Wang T, Zhu Y, Getahun Z, Du D, Huang CY, Degrado WF, et al. Length Dependent Helix-Coil Transition Kinetics of Nine Alanine-Based Peptides. J Phys Chem B. 2004;108.
[301] Gornall JL, Terentjev EM. Universal kinetics of helix-coil transition in gelatin. Physical review E, Statistical, nonlinear, and soft matter physics. 2008;77:031908.
[302] Sterpone F, Derreumaux P, Melchionna S. Protein Simulations in Fluids: Coupling the OPEP Coarse-Grained Force Field with Hydrodynamics. J Chem Theory Comput. 2015;11:1843-53.
[303] Chiricotto M, Melchionna S, Derreumaux P, Sterpone F. Hydrodynamic effects on beta-amyloid (16-22) peptide aggregation. J Chem Phys. 2016;145.
[304] Eliezer D. Biophysical characterization of intrinsically disordered proteins. Curr Opin Struc Biol. 2009;19:23-30.
[305] Zhang WH, Chen JH. Efficiency of Adaptive Temperature-Based Replica Exchange for Sampling Large-Scale Protein Conformational Transitions. J Chem Theory Comput. 2013;9:2849-56.
[306] Zhang WH, Chen JH. Replica Exchange with Guided Annealing for Accelerated Sampling of Disordered Protein Conformations. Journal of computational chemistry. 2014;35:1682-9.