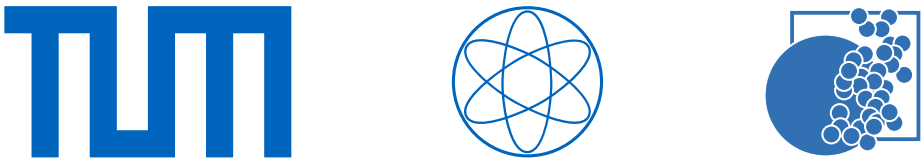
PHYSIK–DEPARTMENT

TECHNISCHE UNIVERSITÄT MÜNCHEN

Dissertation

Structure and Dynamics of Intrinsically Disordered Proteins upon Phosphorylation

Rainer Bomblies



PHYSIK–DEPARTMENT

TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Theoretische Biophysik (T38)

Structure and Dynamics of Intrinsically Disordered Proteins upon Phosphorylation

Rainer Bomblies

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:	Prof. Dr. Hendrik Dietz
Prüfer der Dissertation:	1. Prof. Dr. Martin Zacharias
	2. Prof. Dr. Iris Antes

Die Dissertation wurde am 15.12.2016 bei der Technischen Universität München eingereicht und durch die Fakultät der Physik am 02.05.2017 angenommen.

Wir suchen niemals die Dinge, sondern das Suchen nach ihnen.
Blaise Pascal (1623 - 1662)

Acknowledgments

This thesis is the result of the four years of my time as a PhD student, and none of it would have happened without the tremendous help and support I received from so many people along the way. So allow me a moment to thank all the people that made it possible for me to accomplish the fascinating project that my PhD turned out to be. First and foremost, my thanks go to my supervisor Martin Zacharias, best of the best, who gave me the opportunity to dive into the world of theoretical biophysics. Even when entering with worries, I always left his ever open office energized and with great new ideas. I also want to thank the examination board, Prof. Prof. Dr. Iris Antesfor the examination report and Prof. Dr. Hendrik Dietzfor chairing the examination. I am grateful to all our collaborators, the Technische Universität München (TUM), Deutsche Forschungsgemeinschaft (DFG) with their Sonderforschungsbereich (SFB) 1035 and the Leibniz Rechenzentrum (LRZ) for facilitating the research that resulted in this thesis. I thank all the people who over the years were part of our T38 group and made this a unique place to work and wonder about biophysics. My thanks go to Manu, I will always remember the great times in all the "offices" we shared from Garching to Baltimore and am grateful to have found a dear friend. Thanks to Fabi for all the hours we spent on fascinating new projects and premature celebrations. Thanks to the ever warm and welcoming Sonja for taking care of so many things in my best interest. And thanks to Piotr, Mahmut, Giuseppe, Alex, Flo, Christina and all the others from T38 for many happy hours spent with physics and other stories.

Personally, now that my long time as a scholar finally comes to an end, I want to thank all my friends, who made me who I am today and accompanied me on the many parts of the journey, from the glorious days of the Schollheim via Padua and Edinburgh to Schwabing. My deepest gratitude goes to my family, to my grandparents, my aunts and uncles, parents-, sisters- and brothers-in-law for all the attention and support and to my parents and siblings for being the foundation that I know will always be there for me. My final thanks go to my son Raphael for allowing me at least some sleep and providing that little bit of extra motivation, and to my wonderful wife Anne for the love and joy you keep bringing into my life every single day.

Abstract

The description of intrinsically disordered proteins (IDPs) has challenged the traditional dogma of biophysics that proteins need a defined three-dimensional structure for their biological function. The flexibility and accessibility of such disordered regions makes them a prime target for post-translational modifications like phosphorylation. Control of function of IDPs via phosphorylation has been identified as a key factor in several signaling pathways. The exact effects of phosphorylations on IDPs are, however, still unknown. In this work, molecular dynamics (MD) simulations with their precise temporal and spatial resolution and advanced sampling techniques were used on several systems involving intrinsic disorder. The performance of current MD force fields with respect to IDPs was compared to experimental NMR results. Results show, that disordered behavior is consistently reproduced correctly. The exact content of residual structure, however, varies depending on the choice of the force field. The binding of the protein domains pKID and KIX is an example of an IDP (pKID) that, upon phosphorylation and binding to KIX, undergoes a transition to a folded structure. This work reports of simulations in which, already in the unbound pKID, phosphorylation induces a shift of the conformational equilibrium towards the bound structure. Further simulations reveal important interactions and a possible association pathway, where an initial contact of the phosphorylation site leads to successful folding and binding. The IDP 4E-BP2 only upon phosphorylation folds into a stable structure that prevents signal transduction. Simulations conducted for this thesis revealed both local and global stabilizing effects of the phosphorylations and discovered interactions that are essential for the stability of the fold. This thesis provides evidence that, while there is still room for improvement, simulations can contribute significantly to the investigation of IDPs and their reactions to phosphorylation.

Zusammenfassung

Ein traditionelles Dogma der Biophysik, die Notwendigkeit einer definierten dreidimensionalen Struktur für die biologische Funktion von Proteinen, wurde um die Jahrtausendwende durch die Beschreibung intrinsisch ungeordneter Proteine (IUPs) in Frage gestellt. Mit ihrer Flexibilität und guten Zugänglichkeit sind solche ungeordneten Regionen eines der wichtigsten Ziele für post-translationale Modifikationen wie Phosphorylierungen, die ein Schlüsselfaktor für die Funktionalität von IUPs in verschiedenen Signalwegen sind. Die genauen Effekte von Phosphorylierungen auf IUPs sind jedoch bis heute meist unbekannt. Diese Arbeit nutzt Molekulardynamik-Simulationen (MD) mit ihrer präzisen räumlichen und zeitlichen Auflösung und verbesserte Sampling-Methoden für die Untersuchung von Systemen mit intrinsischer Unordnung. Die Leistungsfähigkeit aktueller MD Kraftfelder bezüglich ungeordneter Proteine wurde mit experimentellen NMR Ergebnissen verglichen. Hierbei reproduzierten alle Kraftfelder intrinsische Unordnung zuverlässig, die Häufigkeit von Sekundärstrukturen hing jedoch von der Wahl des Kraftfelds ab. Die Assoziation der Proteindomänen pKID und KIX ist ein Beispiel eines IUPs (pKID), das nach der Phosphorylierung und der Bindung an KIX eine gefaltete Struktur einnimmt. Diese Arbeit beschreibt Simulationen in denen die Phosphorylierung schon bei ungebundenem pKID das Konformationsequilibrium in Richtung der gefalteten Struktur verschiebt. Weitere Simulationen beschreiben und quantifizieren wichtige Interaktionen zwischen den Bindepartnern und einen möglichen Assoziationspfad bei dem ein anfänglicher Kontakt der Phosphorylierungsstelle zum erfolgreichen Binde- und Faltungsprozess führt. Das IUP 4E-BP2 formt nur nach Phosphorylierung eine stabile Struktur, die die Signalübertragung verhindert. Simulationen im Rahmen dieser Arbeit identifizierten sowohl lokale als auch globale stabilisierende Effekte der Phos-

phorylierungen und deckten Wechselwirkungen auf, die unverzichtbar für die Stabilität der Faltung sind. Diese Dissertation zeigt, dass bei allem Raum für Verbesserungen Simulationen schon heute signifikant zur Untersuchung von IUPs und deren Reaktionen auf Phosphorylierung beitragen können.

Contents

1. Introduction

Whenever mankind hoped to have understood some of the mechanisms of nature the reality proved to be far more complicated and harder to grasp. Proteins are thought of as the machinery of biology, and this image can be quite helpful for understanding biological mechanisms. But the properties of these biomachines can be vastly different to the typical machines made by men. Proteins are not necessarily rigid, they do not necessarily possess a 3D structure and can still exert their function perfectly, sometimes even better. The discovery of these intrinsically disordered proteins (IDPs) around the turn of the millennium has opened up a whole new field of study for biophysics.

IDPs are difficult to study with the experimental methods available to biologists, chemists and physicists. Most techniques rely on ensemble averages that, in the case of IDPs, can not resolve individual substates of the ensemble of conformational states. The high temporal and spatial resolution of computational models like molecular dynamics simulations can in this case be of great value. For small systems straightforward simulations of the system can reach the time scales necessary to cover all relevant parts of the phase space in sufficient detail. For most systems, however, simulations with currently available hardware require prohibitively long computation times and advanced sampling techniques have to be devised and improved to access the full phase space. It is the aim of this study to investigate the influence of post-translational modifications on intrinsically disordered proteins. To this end classical, continuous MD simulations are performed, but also advanced sampling methods are adjusted to allow sampling of the complete conformational space of the segments of interest.

This thesis is structured as follows:

- This first chapter gives an introduction to intrinsically disordered proteins and post-translational modifications.
- In the second chapter the basic principles of molecular dynamics simulations are explained and the problems with IDPs are addressed.
- The third chapter compares the performance of water models specifically designed for IDPs to a traditional water model.
- In chapter four the effects of phosphorylation and coupled folding and binding of the intrinsically disordered domain pKID to the binding partner KIX are investigated both with continuous MD simulations and advanced sampling methods.
- Chapter five examines the reasons of structure stabilization upon phosphorylation of the 4E-BP2 protein with several methods.

1.1. Intrinsically Disordered Proteins

During the last century the field of structural biology was more and more convinced of Nobel prize laureate Anfinsen's dogma[1], that the amino acid sequence determines the three-dimensional structure of proteins. In other words the dogma claims that all proteins possess a unique, stable and kinetically accessible global free energy minimum. This defined structure was assumed to be necessary for the proteins to exert their biological function.

In the last decades, however, evidence has accumulated, that not all proteins are in one single, folded state at all times. Around the turn of the millennium, the previous, scattered experimental observations of disordered proteins lead to a number of publications starting to question the necessity of a stable fold for the function of all proteins.[2–6] These seminal papers introduced the concept of intrinsically disordered regions (IDRs) or intrinsically disordered proteins (IDPs) that do not possess a defined structure but still exert a biological function.

The single unifying property of IDPs is their lack of a unique and stable tertiary structure. Initially, IDPs were termed intrinsically unstructured, but with increasing knowledge about the forms of intrinsic disorder the term 'unstructured' turned out to be more and more misleading. IDPs do not necessarily lack any kind of structure, but can be roughly classified in three major forms of intrinsic disorder: Collapsed disorder, semi-collapsed disorder and extended disorder (see Figure 1.1). Collapsed disordered proteins[5,7,8] do form secondary structure elements that can move as a unit and fail to remain in stable tertiary structures. Side chains of these proteins are typically flexible. Semi-collapsed disordered proteins form due to water being a poor solvent for the peptide backbone. They can show quickly changing H-bonds between side chains and include pre-molten globules[8], polyglutamine regions[9] and other polar sequences[10]. Pre-molten globules, in particular, can still show secondary structure elements. Finally, proteins in extended disorder come closest to the classic idea of a random coil[7,8]. Their disorder is often induced by high net charges and a lack of hydrophobic residues or a hydrophobic core.

The amino acid sequence, although not necessarily determining one defined structure of a protein, does contain information on the folding properties of the

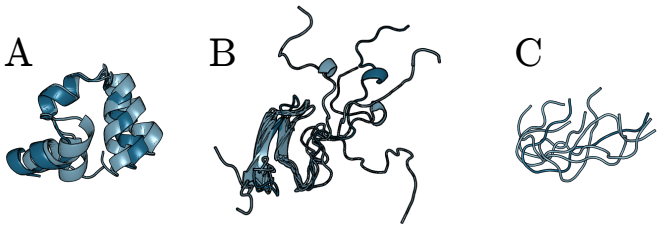


Figure 1.1.: Examples for the three major forms of disorder: Collapsed disorder (A), semi-collapsed disorder (B) or extended disorder (C).

protein. Intrinsically disordered regions typically contain few hydrophobic side chains and increased numbers of charged side chains[5]. Specifically developed sequence predictors[11] use machine learning algorithms to estimate the level of order or disorder from the sequence. These estimates find structural disorder to be not only present but abundant in all species and kingdoms of life, but more so in eukaryotes compared to prokaryotes[12]. It is estimated that 10–35% of prokaryotic and about 15–45% of eukaryotic proteins have segments of significant disorder[13], i.e. disordered stretches of 30 or more consecutive residues. As a tendency, disorder increases with complexity. The highest level of disorder is, however, not found in the most complex metazoan eukaryotes like humans, but in single-celled eukaryotes.[12]

How Can We Learn About Disordered Proteins?

The disordered nature of proteins or regions can be traced with several experimental methods. X-ray crystallography or nucleo-magnetic resonance (NMR) experiments can give structural insights. Circular dichroism (CD) spectroscopy can identify – possibly transient – secondary structure content. Information on the single molecule level can be obtained from Förster-resonance energy transfer (FRET), fluorescence correlation spectroscopy (FCS) or atomic force microscopy (AFM). Finally, mass spectrometry, small angle X-ray scattering (SAXS) or protease digestion can reveal global, average properties.

NMR measurements are probably the most common technique to investigate IDPs.[13–16] Secondary structure can be identified via chemical shifts of the backbone[17] or with information about backbone dihedral angles from scalar couplings[18]. Secondary or tertiary structure can be inferred from nuclear

Overhauser effect (NOE) experiments containing information on spatially close residues or residual dipolar couplings (RDCs)[19,20] reporting on the angle of a bond relative to an external magnetic field. For long range restraints paramagnetic relaxation enhancement experiments (PRE) have proven useful[21], and relaxation rates can provide data on the dynamics of IDPs[14]. X-ray crystallography can identify intrinsically disordered proteins or regions through their missing electron density.[5] However, whether the missing electron density stems from disorder or is an artifact of the purification process or measurement must be tested by a different method, and no insight about possible transient structure can be obtained. The scattering intensities of SAXS give ensemble-averaged estimates of the shape and radius of gyration of proteins[22,23]. Large radii of gyration for a given molecular weight, for example, point towards a unfolded structure. Typically, SAXS measurements complement other measurements like NMR. The differential absorption of left- and right-handed circularly polarized light used in CD spectroscopy contains information on secondary structure content[24] and can semi-qualitatively tell if proteins contain secondary structure or if unstructured proteins fold when in contact with binding partners[25]. In contrast to the methods presented so far single molecule studies on IDPs do not average over a large sample[26]. Single molecule FRET[27,28] can measure the distance between two chromophores using the Förster energy transfer distance dependence and can help in classifying conformations or identifying long-range interactions between domains. Using similar chromophores, FCS allows the calculation of the autocorrelation function of the distance of the chromophores and can be applied to IDPs to check, for example, their flexibility[29] or hydrodynamic radii[9]. AFM can mechanically measure changes of proteins like unfolding and can assess conformational equilibria[7] or even monitor conformational changes[30]. Finally, mass spectrometry[31], the digestion by protease[32] and many other experimental techniques have also been successfully applied to study IDPs.

Simulations, in particular molecular dynamics (MD) simulations, also significantly contribute to the investigation of IDPs. Already at the level of structure determination via NMR structural biologists typically use simulations. These simulations narrow down possible structures via simulated annealing simu-

lations or torsion angle dynamics[33,34], where NMR results are translated to restraints for short simulations scanning the conformational space.

Additionally, several hybrid experiment-simulation approaches have been developed to characterize the conformational ensembles of IDPs.[35,36] The ENSEMBLE algorithm[18], as an example, includes a variety of experimental restraints into simulations to determine the simplest population weighted ensemble of structures that reproduces the experimental data. Experimental input from chemical shifts, NOEs, PREs, residual dipolar couplings, hydrogen exchange protection factors, solvent-accessible surface area, and the hydrodynamic radius can be used. However, when using simulations with experimental restraints, care must be taken to reproduce the experimental conditions like temperature, pH and pressure[37], otherwise an unphysical ensemble might be generated.

Simulations can work well with experimental data but can also contribute *ab initio*.[38] With atomic detail and full dynamics up to femtosecond resolution MD simulations are, in principle, ideal for the investigation of the behavior of IDPs. The purely theoretical simulation predictions are, of course, most valuable if they are validated with experiments[39]. A prime example of such a validation is a simulation study on Aβ40 and Aβ42[40] that confirmed the predicted ensemble by calculating predicted scalar couplings that were in quantitative agreement with NMR measurements. For disordered entropic chains there is, currently, no high-resolution structural information available from experiments. For such systems, *ab initio* MD simulations have proven to be very useful, like in the case of elastin[41]. The level of detail in IDP simulations ranges from explicit solvent[40] through implicit solvent, like the ABSINTH model[42], specifically developed for IDPs, to coarse-grained models[43] that do not represent every atom explicitly. The computational cost of the calculations, however, increases with the level of detail and thus can limit the applicability of simulations. Details on the problems with simulations of IDPs will be covered in the theory part of this work.

Is Disorder Relevant?

Disorder has been discovered and mostly studied in the artificial and dilute environment *in vitro*, but are IDPs actually disordered *in vivo*? Experiments mimicking the macromolecular crowding of a live cell with a 400 mg/ml concentration of Dextran or Ficoll[44,45] showed, that disorder persists in crowded environments. Increased digestion of IDPs by the ubiquitin-independent 20S Protease[46] was observed both *in vitro* and *in vivo*, indicating the consistent disordered structure that is more easily cleaved by the protease. Furthermore, NMR studies on isotope-labeled proteins demonstrated their intrinsic disorder in living cells.[47,48]

Many human diseases are associated with disordered proteins or regions.[49] Analysis of the SwissProt data bank revealed that many diseases, including cancer, malaria, human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS), deafness, obesity, cardiovascular diseases, diabetes mellitus, albinism, and prion protein related diseases, are correlated with proteins predicted to be disordered[49,50]. Keith Dunker and colleagues[51] in 2002 found, that 79% of cancer-associated and 66% of cell-signaling proteins contained regions of predicted disorder that were 30 residues or longer. Of a comparable set of the eukaryotic proteins of the SwissProt database only 46% contained such long regions of predicted disorder. One example for the importance of disorder in diseases is the transcription factor p53[52], regulating cell cycle progression, apoptosis induction, DNA repair, and response to cellular stress. Of the many interactions of p53 approximately 70% are mediated by IDRs[53]. Mutations of p53 are linked with cancer in colon, lung, esophagus, breast, liver, brain, reticuloendothelial and hemopoietic tissues[54]. Another example is Aβ protein, which under physiological conditions is a random extended chain and shows no α-helical or β-sheet like conformation[55]. The protein can, however, aggregate and form fibrils closely linked with Alzheimer's disease and correlating with neurotoxicity.[56]

What Is the Function of Disorder?

Intrinsically disordered proteins play an important role in transcription, translation, cell cycle regulation and cell signaling[57–61], as indicated by their higher abundance in the eukaryotic proteome with its extensive interaction network compared to the prokaryotic proteome[12]. When interacting with other cellular components, IDPs often show disorder to order transitions.[62] Their high specificity and low affinity binding[58] is ideally suited for signal transduction, as the high specificity guarantees precise signal delivery while the low affinity allows fast dissociation after signaling is complete. Post-translational modifications (PTMs), also highly important for signal transduction, are often associated with IDPs/IDRs[63]. The flexibility of IDPs allows specific binding both to the physiological target and to modifying enzymes[58]. Furthermore, intrinsic disorder promotes digestion by proteases[5], which could be beneficial for signal transport as IDPs would have a shorter life cycle due to faster degradation[3]. Finally, IDPs can bind faster to binding partners due to a "fly-casting" effect.[64,65] The unstructured protein can have a greater capture radius for a specific binding site compared to a folded protein[64]. In the fly-casting scenario, the unfolded state then binds weakly and only folds at the binding site. The effect would be particularly beneficial for hub proteins in protein-protein-interaction networks, where indeed intrinsic disorder is commonly found[66]. Another possible explanation for the faster association rate is based on the fact, that coupled folding and binding shows a reduced free energy barrier of binding that accelerates binding[27]. According to this view, the capture rate of IDPs is even reduced compared to folded proteins due to the higher diffusion constant, but fewer encounters are necessary for successful binding.

The function of IDPs is a good means for classification, although many other classifications, like their functional elements, their structure or sequence, their interactions with other proteins, their evolution, regulation or biophysical properties[67] could also be used. One possible and exemplary classification by function is in six classes: Entropic chains, display sites, chaperones, effectors, assemblers, and scavengers[6,67,68]. Entropic chains directly benefit from their disorder, for example as flexible linkers or as spacers between regions. IDPs contain a disproportionate, increased number[69] of PTM sites that can affect

properties of the protein. These modification sites are more easily accessible in disordered regions[70] and thus IDPs can act as display sites for possible PTMs. IDPs also act as chaperones helping RNA or other proteins to fold correctly. Here, intrinsic disorder can be beneficial to prolong the life time of encounter complexes by sampling several possible encounter conformations. The class of effectors contains IDPs that interact with other proteins and modify their activity. Often this is accompanied by disorder-to-order transitions.[71] Next, assembler IDPs bind multiple proteins and through binding either stabilize their complex or locally confine them, promoting interaction between the bound proteins[72]. The final class are scavengers that store or neutralize small ligands, like casein binding calcium phosphate in milk[73].

IDPs exert their function through one of three classes of functional motifs.[67] The first class are short linear motifs (SLiMs) of 3-10 amino acids in length[74,75]. Interactions with SLiMs can either modify the motifs themselves or the motifs can act as recognition peptides. The motifs can be cleaved by proteases[76], modified by enzymes for PTM addition or removal[77], or undergo structural changes, like cis–trans isomerization of a proline through Pin1[77]. When acting as recognition peptides, the motifs can either be used for complex formation[78], for docking by increasing specificity and efficiency of protein modifications[79], or for targeting proteins for transport towards subcellular organelles. Importin, for example, recognizes the nuclear localization signal motif and transports the respective proteins to the nucleus.[80] The second class of function, termed molecular recognition features (MoRFs)[81], are domains of 10-70 amino acids that promote specific protein–protein interactions. They typically undergo disorder-to-order transitions when binding to partner proteins[82] and the unbound, disordered form can be biased towards the bound, folded form[83]. Finally, the third functional motif are permanently disordered domains. Some domain families seem to solely require the presence of disorder in their local environment for functioning, while others seem to rely on the presence of disordered regions in specific locations[84,85].

Among the many functions of disordered proteins and regions is their propensity to host PTM sites. The next section is intended to give a broad overview on the ubiquitous regulation mechanisms of PTMs.

1.2. Post-translational Modifications

The complexity of intrinsic disorder is only one of the many open questions about the biological machinery of the proteome. The enormous diversification of proteins in form and function is another. The initial coding capacity of the DNA is, in the course of transcription and translation, expanded to generate an immense diversity in the proteome. For example, the roughly 30,000 genes in the human genome are the blueprint for more than a million molecular species of proteins. Two major mechanisms allow this expansion. The first one acts act transcriptional level, where through splicing of the mRNA, the working copy of the DNA, the actually expressed exon is created[86]. The second route to diversification, when the ribosome has already translated the mRNA into the polypeptide chain of the protein, is called post-translational modification (PTM). As many as 300 different PTMs of proteins are known to occur physiologically.[87] They include modifications at any time of a protein's life cycle, from shortly after translation where folding can be assisted, to ubiquitination as a marker for degradation[88].

PTMs are covalent modifications either to the peptide linkages of the amino acid residues or to their side chains.[89] This definition already separates the PTMs in two broad classes: The first are cleavages of the covalent peptide backbone bond either by proteases or, less frequently, by autocatalytic cleavage. Selective proteolysis is a fundamental mechanism to regulate the composition of the proteome and to control location, activity and lifetime of proteins both inside and outside the cell. In eukaryotic cells, large subsets of proteins are subject to limited proteolytic clipping as part of the normal spatial and temporal protein maturation process. As an example, essentially every protein that enters the endoplasmatic reticulum (ER) loses 25-30 C-terminal residues through proteolysis, since the cut tail was used as signal sequence specifying transport to the ER[90].

The second class includes all covalent addition or removal of chemical groups catalyzed by enzymes. Typically, an electrophilic fragment of a co-substrate is added to a side chain. The side chain is usually electron rich and acts as a nucleophile in the transfer.[89] PTMs are believed to increase the number of chemically different amino acids in proteins from 20 to more than 140[89].

Important Forms of Covalent PTM Additions

The five most prevalent covalent attachment modifications are phosphorylation, acylation, alkylation, glycosylation, and oxidation.[89] Oxidation plays an important role when two thioles are merged to a disulfide bond that stabilizes the protein fold. Glycosylation is the attachment of sugar molecules. The most common form is N-glycosylation of Asparagine side chains in the ER that is crucial for glycoprotein folding[91]. Alkylation as methylation also takes part in histone-coding but can also add long isoprenyl chains for localization at membranes[89]. Acylation most commonly happens with C_2, C_{14} or C_{16} chains or as covalent attachment of ubiquitin, with vastly different effects. C_2 acetylation patterns of the Lysine residues of the N-termini of histones can regulate the transcription machinery through the 'histone-code'[92]. Addition of C_{14} or C_{16} fatty acyl groups can direct proteins to the membrane[93]. Ubiquitination, which leads to degredation of a protein, is enzymatically transferred as an acyl moiety and thus is also categorized as acylation[94].

Finally and most commonly, phosphorylation substitutes a hydroxyl group with a charged, dianionic tetrahedral phosphate group (see Figure 1.2). This change is catalysed by dedicated phosporylation proteins, the kinases, exploiting ATP as the universal phosphate donor. Their superfamily, the kinome, has more than 500 members in the human genome[95]. The mammalian phosphoproteome contains phosphoserine(pS), phosphothreonine(pT) and phosphotyrosine(pY) residues at a ratio of approximately 90(pT):10(pT,pY)[96]. Bacterial and funghal proteome may also contain phosphohistidine and phosphoaspartate[97]. The conversion of a neutral OH group to a dianionic phosphate group is a potent conformational switch as it has evolved into a major recurring pattern to create proteome diversity in eukaryotes.

Phosphorylation, but also acetylation or acylation, are reversible via deconjugating enzymes. Modifying proteins facilitates a fast and economical control of protein function, as degradation and de novo synthesis would be energetically far more costly and much slower. These advantages have lead to a dominant role of phosphorylation for protein based signaling in eukaryotes.[89] The cyclic AMP activated kinase A for example, phosphorylates over 100 proteins[98] on Ser and Thr side chains for signal transduction. To turn off signals the phos-

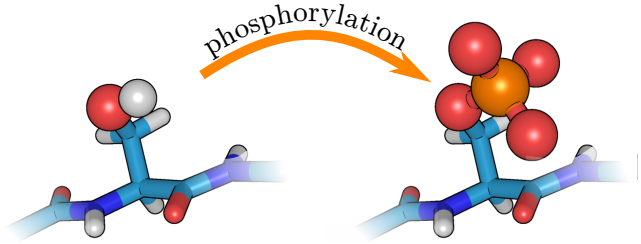


Figure 1.2.: Phosphorylation of a serine residue.

phate group has to be removed by phosphatase enzymes, of which there are approximately 150 encoded in the human genome. Interestingly, most of these, (>100) are phosphatases for the less frequent pY phosphorylation[99], as the pS/pT phosphatases are typically promiscuous and contain many regulatory subunits that control subcellular location and substrate recognition.

The mechanisms by which phosphorylation alters the properties of a protein are diverse. Phosphorylations can prevent activity of an enzyme by sterically blocking the active site[100] or act as a recognition site for binding proteins[101]. In particular, however, the special properties of phosphorylations can influence the equilibrium of structures of the protein. At physiological pH the phosphate group with a pK_A of ~ 6.7 is predominantly dianionic[102]. The two negative charges and, due to the three phosphoryl oxygens, the possibility to form several hydrogen bonds constitute very special characteristics. Three direct interaction types for structural control are observed.[103] Firstly, the possibility to form several H-bonds is used in tight environments to contact other, mostly arginine, side chains. Secondly, the negative charge can interact with the positive charge of the N-terminus of a helix and the phosphate group can interact with the backbone nitrogen atoms of the helix. Thirdly, the phosphorylation can be involved in contacts to other polar residues like lysine or histidine. Through these interactions phosphorylations have been found to induce disorder[104] and order[105–108], in proteins.

It is the aim of this work to use molecular dynamics simulations in order to elucidate some effects of phosphorylation on intrinsically disordered proteins for two test systems where phosphorylation stabilizes folded structures. In the next chapter, a brief introduction to the methods of MD simulations will be given.

2. Theory

The fundamental task of theoretical physics is to rationalize, explain and predict phenomena occurring in the world we live in. In theoretical biophysics, in order to rationalize, explain and predict the mechanisms and behavior of biomolecules, models of these molecules have to be built that consider all, but no more than the relevant interactions and degrees of freedom. In this work, molecular dynamics (MD) computer simulations were used, which will be described in the following section.

2.1. Molecular Dynamics Simulations

Molecular dynamics simulations describe the dynamics of a system of interest by modeling every atom as a point adhering to the classical Newtonian equations of motion:

$$m_i \frac{d^2 r_i(t)}{dt^2} = F_i \tag{2.1}$$

where the force F_i acts on atom i with mass m_i. The model of atoms following classical mechanics is based on the Born-Oppenheimer approximation, where nuclei with their 2000-fold higher mass are assumed to only feel average forces from their faster, surrounding electrons. The approximation turned out to be a very good for a wide range of materials and in particular for biomolecules[109]. Forces are calculated from an empirical potential function V termed force field [1].

$$F_i = \frac{\partial V(\mathbf{r}_1, ..., \mathbf{r}_N)}{\partial r_i} \tag{2.2}$$

[1] 'Force field' is henceforth used in this chemical meaning as a scalar potential and not as the physical term, a vector field of forces.

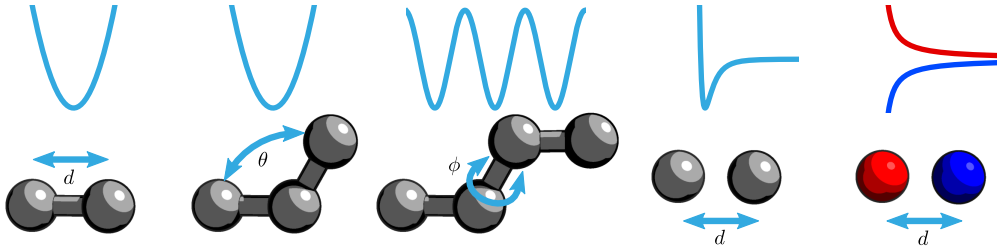


Figure 2.1.: Contributing terms to the force field and the typical shapes of their potential. The first three terms describe bonded interactions while the last two, Van-der-Waals and Coulomb interactions, apply for non-bonded atoms.

Since the physics of the individual molecules is coarsened to complete atoms, the model is stripped of the original source of forces acting on the atoms. Formally, atoms in biomolecular systems are subject to forces from covalent bonds stemming from shared electronic orbitals, Van-der-Waals interactions between induced dipoles of the atoms and Coulomb interactions. Coulombic interactions are, in principle, straight forward to implement and the contribution to the potential is simply the Coulomb potential. For the other two contributions, however, coarsening the representation deprives us of the possibility to calculate overlaps between electronic shells or induced dipole moments in the atoms. So, to correctly represent the forces acting on an atom, these effects have to be included in artificial force field terms. The Van-der-Waals interactions include the repulsion of an overlap of the electronic shells and a weak attraction due to induced dipoles. With a strong repulsion at close distances and a slight attraction beyond, these interactions are most commonly represented with the Lennard-Jones potential[110]. Covalent contributions to the potential in MD are represented by three terms: Bonds and angles between atoms are implemented as harmonic springs, while for dihedral terms a periodic potential with local minima is used. In total, the potential function thus contains five contributing terms, depicted in Figure 2.1, and has the form of equation 2.3.

$$
\begin{aligned}
V = & \sum_{\text{bonds}} k_i^{\text{bond}}(d_i - d_{i,0}) \\
& + \sum_{\text{angles}} k_i^{\text{angle}}(\theta_i - \theta_{i,0}) \\
& + \sum_{\text{dihedrals}} k_i^{\text{dihedral}}(cos(n\phi_i + \delta_i)) \\
& + \sum_{i,j \neq i} 4\epsilon \left(\left(\frac{\sigma}{d_{ij}} \right)^{12} - \left(\frac{\sigma}{d_{ij}} \right)^{6} \right) \\
& + \sum_{i,j \neq i} \frac{q_i q_j}{\epsilon d_{ij}}
\end{aligned}
\tag{2.3}
$$

Coordinates of all atoms and thus the distances between atoms d_i, angles θ_i and dihedrals ϕ_i as well as the distances d_{ij} between non-bonded atoms are known. All remaining parameters have to be defined empirically[111], by fitting to quantum-mechanical calculations and experimental, e.g. spectroscopic, data. Several versions of such force fields have been implemented. The most prominent today are the AMBER[112], CHARMM[113] and GROMOS[114] force fields.

With all necessary parameters for the force field defined, the forces on atoms for a given set of atom coordinates can be calculated. To obtain the dynamics of the system, these forces have to be translated into atom motion. This is achieved by integrating Newton's equations of motion of Equation 2.1. For this task several integrators have been developed. The Verlet algorithm[115] expands the position $\mathbf{r}$ of each atom in a Taylor expansion for a short time Δt before and after a certain time t:

$$
\begin{aligned}
\mathbf{r}(t + \Delta t) &= \mathbf{r}(t) - \frac{d\mathbf{r}(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2\mathbf{r}(t)}{dt^2}(\Delta t)^2 - \frac{1}{6}\frac{d^3\mathbf{r}(t)}{dt^3}(\Delta t)^3 + \mathcal{O}(\Delta t)^4 \\
\mathbf{r}(t - \Delta t) &= \mathbf{r}(t) + \frac{d\mathbf{r}(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2\mathbf{r}(t)}{dt^2}(\Delta t)^2 + \frac{1}{6}\frac{d^3\mathbf{r}(t)}{dt^3}(\Delta t)^3 + \mathcal{O}(\Delta t)^4
\end{aligned}
\tag{2.4}
$$

Upon addition of the two equations the uneven exponents cancel out and the position at a later time $t + \Delta t$ is only dependent on previous positions and the

acceleration that can be obtained from the potential and the mass of the atom:

$$\mathbf{r}(t+\Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\Delta t) + \frac{d^2\mathbf{r}(t)}{dt^2}(\Delta t)^2 + \mathcal{O}(\Delta t)^4 \tag{2.5}$$

The Verlet algorithm requires two initial sets of coordinates. This is solved in the Velocity Verlet[116] algorithm, where it is sufficient to know one initial set of coordinates with respective velocities. As velocities are necessary for the computation of global statistical properties, their calculation does not expand computational demand. Update rules for the Velocity Verlet algorithm are

$$\begin{aligned} \mathbf{r}(t+\Delta t) &= \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)(\Delta t)^2 \\ v(t+\Delta t) &= \mathbf{v}(t) + \frac{\mathbf{a}(t) + \mathbf{a}(t+\Delta t)}{2}\Delta t \end{aligned} \tag{2.6}$$

As an alternative, the leap frog algorithm[117] by turns calculates velocities and positions at half steps to obtain the trajectory of positions:

$$\begin{aligned} \mathbf{r}(t+\Delta t) &= \mathbf{r}(t) + v(t+\Delta t/2)\Delta t \\ \mathbf{v}(t+\Delta t/2) &= \mathbf{v}(t-\Delta t/2) + \mathbf{a}(t)\Delta t \end{aligned} \tag{2.7}$$

For all integrators, the size of the time step Δt is vital. On the one hand, long simulation times are desirable to generate a large ensemble, so Δt should be as large as possible. On the other hand, positions of the atoms cannot be resolved correctly when the time step is larger than the fastest vibrations in the system. In biophysics, these are typically vibrations of hydrogen atoms with a frequency of 10 fs, so the time step of MD simulations has an upper limit of 1 - 2 fs. Algorithms to satisfy bond geometry constraints like SHAKE[118] and the repartitioning of hydrogen masses[119] can help increase the time step so that simulations with a time step of 4 fs can be performed.

Integrators can only propagate coordinates according to acting forces but need an initial set of coordinates to start with. For simple systems, like ions in a box of water or short peptides, these initial coordinates can be generated *ab initio*. For larger systems, like proteins or DNA, a structure with atomistic resolution from NMR or X-ray crystallography measurements has to be used.

Starting coordinates complete the prerequisites and the MD simulation can be performed as a cycle of consecutive force calculations and position (and velocity) updates. When the desired length of the simulation is reached, the cycle is stopped and the trajectory, i.e. the coordinates of all atoms saved at a given time interval, is complete and ready for analysis.

2.1.1. Statistical Ensembles in MD

The ergodic hypothesis states, that for a long time span the probability of a state in an ensemble is proportional to the time spent in this state by the system. Assuming this hypothesis, plain integration of Newton's equation of motion samples the microcanonical ensemble, albeit with a slight energy shift due to numerical rounding errors. Additionally, for the leap frog algorithm the potential and kinetic energies are never known at the exact same time but with half a time step between them. Simulation results should be compared to experiments, where not the system energy, which is constant for the microcanonical ensemble, but rather temperature and pressure are kept fixed. So sampling of the canonical ensemble with the number of particles N, volume of the probe V and temperature T, or the grand canonical ensemble with N, T and pressure p as state variables would be preferable. In order to sample these ensembles, temperature and volume or pressure, respectively, have to be stabilized instead of the total energy.

To stabilize temperature, several methods have been developed. The most simple solution is velocity rescaling, where all velocities of the system are rescaled by a constant factor $\lambda = \sqrt{T(t)/T_0}$, i.e. the square root (as velocities enter kinetic energy calculation quadratically) of the ratio of current temperature $T(t)$ and desired temperature T_0. This forces the system to maintain the correct kinetic energy, but does not allow for fluctuations of the temperature as expected for the canonical ensemble. The Berendsen thermostat[120] allows for fluctuations of the temperature and couples the system to an external heat bath and ensures the change of temperature is proportional to the difference in temperature:

$$\frac{dT}{dt} = \frac{1}{\tau}(T_0 - T(t)) \tag{2.8}$$

The coupling parameter τ governs, how tightly the system is coupled to the heat bath. with $\tau = 1$ simple velocity rescaling would be reproduced. While the Berendsen thermostat does allow fluctuations in temperature it still does not correctly sample the canonical ensemble.

The Nosé-Hoover thermostat[121] allows sampling of the canonical ensemble by expanding the Hamiltonian of the system by an extra degree of freedom for the heat bath, s:

$$\mathcal{H} = \sum_i \frac{\mathbf{p}_i}{2m} + \frac{1}{2}\sum_{i,j\neq i} U(\mathbf{r}_i - \mathbf{r}_j) + \frac{p_s^2}{2Q} + gk_\mathrm{B}T\ln(s) \tag{2.9}$$

where g is the number of independent momentum degrees of freedom and Q is an artificial "mass" that regulates coupling of the system with the heat bath. The additional coordinate s can be interpreted as a scaling factor of the time step and controls the momenta and thus the temperature. As this barostat can in some cases create problems with ergodicity, the Nosé-Hoover formalism can be improved to Nosé-Hoover-Chains[122]. A more detailed discussion of thermostats can be found in[123].

In order to control pressure in simulations, several barostats have been developed. The Berendsen barostat[120] scales the dimensions of the box to produce the correct average pressure, but does not yield the exact NPT ensemble. The Andersen barostat[124], similarly to the Nosé-Hoover thermostat, scales the coordinates of the atoms by introduction of a new term in the Hamiltonian. This was extended in the Parrinello-Rahman barostat[125] so that nonisotropic scaling of the volume is possible.

Equilibration

Starting coordinates generated *ab initio* or taken from crystal or NMR structures can show overlap between atoms or produce other unfavorably large energy contributions due to the force field, which would lead to very high forces and thus velocities and disrupt the system. Also, extreme changes of the temperature or pressure to reach the desired ensemble could lead to unforeseen behavior. So before an actual production MD simulation, several equilibration steps are performed.

First, starting coordinates are brought to a local energy minimum via an energy minimization step with a steepest descent algorithm. Then, velocities are assigned to the atoms according to the Maxwell distribution and temperature of the system is gradually increased. In this phase the kinetic and potential energy contributions balance out. Finally, for simulations in the grand-canonical NPT ensemble, pressure control is included and the system is equilibrated at the desired pressure. The Berendsen thermostat and barostat can quickly converge towards the desired temperature and pressure and thus are often used for equilibration.

With the complete description of the force field, an integrator and equilibrated starting coordinates and velocities production MD simulations can be run. Their results are trajectories that, if the ergodic hypothesis holds, represent the desired statistical ensemble of states of the system. Statistical ensemble averages could also be obtained by Monte Carlo simulations[109], that do not integrate the equations of motion but test a large number of states on their accordance with the Boltzmann distribution of states, typically using the Metropolis criterion[126]. But the trajectories of MD simulations additionally constitute examples of the dynamics,[1] thereby also providing information on the kinetics of the investigated system.

2.1.2. Common Techniques to Facilitate MD Simulations

The relatively small time step compared to the large time scales of the dynamics and the sheer number of atoms in biomolecules limit the applicability of MD simulations. To enhance efficiency of the calculation of MD trajectories, several important methods and ideas have been developed.

Solvent Representation

The first MD simulation of a biomolecule, the simulation of a Trypsin inhibitor by McCammon et al.[127] in 1977, was performed in vacuum. But biological systems never are in vacuum and proteins are not even stable when not in

[1] An interesting discussion on whether or not MD simulations reproduce real trajectories can be found in the textbook of Frenkel[109], pp. 72-74

solution. For a meaningful description the surroundings of the protein also have to be modeled.

One solution is the explicit modeling of solvent atoms in a large box around the molecule(s) of interest. Several water models have been implemented[128–130] and tested. To surround a molecule with sufficient water to reproduce a natural environment, a large number of water molecules is necessary. This can, depending on the required box size, increase the system size by a factor of 10 or more, significantly increasing computational demands. To save computer time, effects of the highly dynamic behavior of the solvent molecules can be averaged in implicit solvent models[131]. The more precise solution to the Poisson-Boltzmann-equation[132] is here still outperformed[133] by the generalized-Born approach[131].

Periodic Boundary Conditions

Free explicit solvent molecules would quickly dissociate far away from the solute of interest. Consequently, all atoms have to be kept in a defined simulation region, the bounding box. This confinement can be achieved via restraints, but then boundary effects at the solvent–vacuum boundary have to be taken into account. To mitigate this boundary problem, and also for further reduction of the size of the simulated box while still mimicking a large system, periodic boundary conditions are used. To this end, copies of the system are placed at all faces of the bounding box, such that a particle that, for example, exits the box to the top would instantly enter the box at the bottom, as shown in Figure 2.2. Only atoms represented by full circles have to be simulated, yet the system feels as if it is solvated in a much larger box, that, however, also contains periodic images of the solute.

Non-bonded Calculations

The large number of atoms makes the quadratic problem of non-bonded interactions practically impossible to solve. A legitimate solution is the use of cut-offs to only calculate interactions with neighboring atoms up to a certain radius. Such a radius can be applied both for Van-der-Waals- and electrostatic

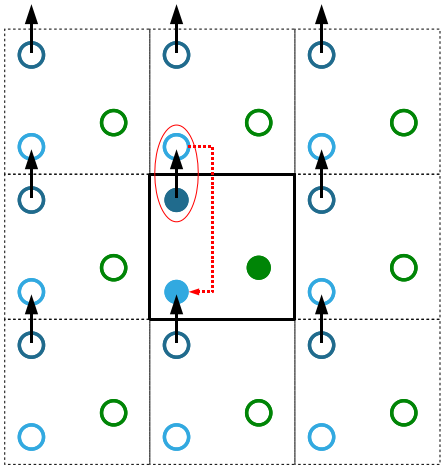


Figure 2.2.: Scheme of periodic boundary conditions. Only full circles are simulated. When the dark blue atom leaves the simulation box a periodic copy is re-introduced as the light blue atom. Figure adapted from [134].

interactions. Artifacts of this cut-off like a steep drop in forces can be softened by shifting the potentials[135] so they smoothly fade to 0 at the cut-off distance.

Neglecting interactions with charges at larger distances introduces a significant error to the calculations, as the Coulomb potential only decreases with $1/d$. To be be precise on close range but still get a good estimate of long range contributions the particle mesh Ewald (PME) method[136] splits the Coulomb potential V_C into two contributions:

$$V_C = \sum_{i,j} V_{C,sr} + \sum_{\mathbf{k}} \mathcal{V}_{C,lr}(\mathbf{k})|\mathbf{æ}(\mathbf{k})|^2 \tag{2.10}$$

The first sum collects the short range electrostatic interactions $V_{C,sr}$ that are calculated from close, pairwise interactions of each *particle*. The second term contains the Fourier transforms $\mathcal{V}_{C,lr}$ of the long range potential and $\mathbf{æ}(\mathbf{k})$ of the charge density, calculated via Fast Fourier transform which evaluates the density on a *mesh* grid. The contributions are summed up as *Ewald* sums[137]. The periodicity required for the Ewald method to work is inherently provided by to periodic boundary conditions.

2.1.3. Analysis of MD Trajectories

The output of an MD simulation, the trajectory, in the ideal case represents the statistical ensemble of states of the simulated system. With the correct ensemble, atomic spatial resolution and, in principle, femtosecond time resolution, many properties of the system can be extracted. Global statistical properties like the temperature or potential and kinetic energies are readily available. But they can also be split into contributions from separate atoms and statistics on the atomic level, like the radial distribution functions of one atom type to another or the root mean square fluctuations of individual atoms or atom groups, can be useful. The dynamics provided by MD can help identify initial contacts for the encounter of ligands or proteins with a binding partner.

A verification of the MD model can be obtained by calculating experimental observables from the MD trajectories. For identified important ensembles algorithms have been developed to predict numerous observables like NMR chemical shifts[138] or SAXS curves[139]. Also, for the dynamics of systems, FRET efficiencies[140] can be predicted.

For the investigation of intrinsically disordered proteins, MD simulations can provide information on the content of order in the IDP, the conformational ensemble, key interactions stabilizing order or invoking disorder or also the time scales of conformational dynamics.

A detailed view on the extraction of free energies from MD simulations shall be given in the following section.

2.2. Free Energies from Molecular Dynamics Simulations

The accurate calculation or prediction of free energies is one of the most important tasks of computational biochemistry. Free energy simulations employ MD or in some cases also Monte Carlo (MC) approaches as techniques to generate appropriate ensembles of states. Within the limitations of the underlying force field, such free energy simulation techniques allow the accurate calculation of free energy differences including both energetic as well as entropic contributions.

As will be further discussed below, a perquisite of accurate free energy simulations is the appropriate sampling of relevant conformational states of a given system, which requires in many cases a series of extensive simulations. Hence, the computational demand of free energy simulations can be enormous. In practice, the application of free energy simulation techniques is largely restricted to absolute free energy calculation of small sets of ligand–receptor complexes or to the calculation of relative free energies associated with chemical modifications of a molecule. It should also be emphasized that current free energy simulations are not truly predictive in the sense that such simulations make the experimental binding affinity determination obsolete[142,143]. In part this can be due to insufficient sampling which limits the convergence of calculated free energy differences. Recent methodological improvements and steadily increasing computer performance, however, have provided progress in this direction. The second important limitation is due to the accuracy of current molecular mechanics force fields used to describe molecular interactions. For proteins and nucleic acids, fairly accurate force fields have been designed in the course of many years. For the many possible organic drug-like molecules that may bind to biomolecules, in contrast, significantly less experience and testing results are available[113,144]. However, even within these limitations numerous useful applications are possible, ranging from qualitative predictions on ligand binding which can help in drug design efforts to offering explanations at the

The contents of this section have been published[141] in similar form.

molecular level for the behavior of a molecule or its alteration due to chemical modification. In this section free energy simulation techniques for alchemical transformations will be reviewed followed by discussing alternative approaches of induced dissociation of the ligand–receptor complexes combined with the calculation of the associated free energy change.

2.2.1. Free Energy in MD Simulations

Over decades the determination of free energies in biomolecular systems has been a major focus of MD simulations. The thermodynamic concept of free energy gives an overall measure for the distribution of available states over a multidimensional energy landscape. A microscopic state in classical MD simulations is thereby given by the Cartesian coordinates for all atoms of the system as well as their momenta. Considering the representation of a biomolecular systems in MD simulations one typically refers to a biomolecule or biomolecular complex which is solvated in a finite sized box of water molecules and surrounding ions. During the simulation the system is kept under constant volume V or pressure p and the temperature T can be controlled with a thermostat representing a thermal heat bath. Such a system corresponds to the canonical or in case of constant pressure to the isobaric ensemble. The Hamiltonian, or energy function, of the system consists of a kinetic energy term and a potential energy function. The potential energy function is represented by a force field description with energy contributions from bonded and non-bonded particle interactions. It usually does not depend on the momenta of the particles so kinetic and position-dependent (configurational) contributions can be separated. In the same way it is also possible to separate kinetic and configurational contributions to the partition function Q and to the free energy F of the system. With $\beta = 1/k_\mathrm{b}T$, k_b being the Boltzmann constant, Q and F can be written as

$$Q = \iint e^{-\beta\mathcal{H}(\mathbf{r},\mathbf{p})}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{p}. \tag{2.11}$$

$$F = -\frac{1}{\beta}\ln Q. \tag{2.12}$$

In common applications for MD simulations on biomolecules it is impossible to explore the phase space and to calculate the complete associated partition function in its entirety due to the high dimensionality of the system. Free energy calculations in practice aim to reduce the need to explore the complete phase space with the help of expressions of free energy differences between two systems A and B:

$$\Delta F = -\frac{1}{\beta} \ln \frac{Q_A}{Q_B} \tag{2.13}$$

The systems A and B can, for example, refer to the receptor with a bound ligand and without a ligand or may correspond to two different conformational regimes of a protein molecule. To determine the difference in free energy between systems A and B it is only necessary to sample those regions in phase space which differ among the systems. In the following paragraph the systems A and B differ chemically and are represented by different force field descriptions.

2.2.2. Alchemical Transformations

Various methods have been suggested to estimate the free energy difference between two systems or distinct system states A and B from MD or MC simulations. In case of free energy simulations on ligand binding one is interested in the absolute binding free energy of a ligand or in the relative binding affinity of two ligands that differ for example in a chemical group. Hence, the states A and B can correspond to a state with ligand absent (A) or present (B) in the receptor binding pocket or in case of considering relative free energy differences the states A or B represent absence or presence of a chemical group in the ligand, respectively. The presence or absence of the ligand or chemical modification can be represented by a difference in the force field description of each system. According to the thermodynamic cycle shown in Figure 2.3 for calculating the contribution to the binding free energy it is necessary to not only calculate the free energy change with the ligand bound to the receptor but also in the unbound state of the ligand (in the bulk solvent).

Zwanzig introduced the free energy perturbation (FEP) formula as an ensem-

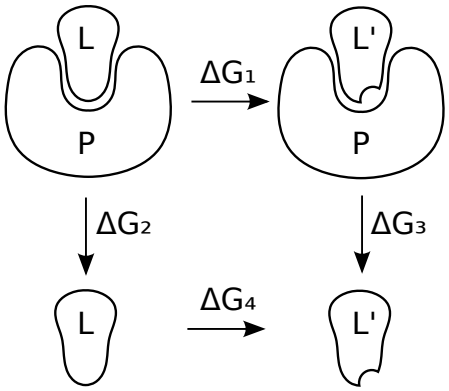


Figure 2.3.: The thermodynamic cycle for calculating the binding free energy change upon ligand modification. The indicated free energy pathways can be calculated from simulations with different approaches. The horizontal pathways (ΔG_1, ΔG_4) involve a change in the system topology as ligand L morphs into L' and are accessible from (alchemical) free energy calculations. Vertical pathways (ΔG_2, ΔG_3) involve a change in the association state indicating the binding event of L and L', respectively, and can be derived from Umbrella Sampling calculations. Typically one is interested in the difference free energy of association which can be obtained from $\Delta\Delta G = \Delta G_1 - \Delta G_4 = \Delta G_2 - \Delta G_3$

ble average over the Boltzmann factor for the energy difference between the two states calculated in the ensemble of A[145]. Hence, it is possible to calculate the free energy change due to a change of the system (a perturbation B, e.g. change in the force field such as an annihilation or addition of a force field term representing a chemical group of the ligand) from the ensemble average of system A (the unperturbed system):

$$\Delta F_{AB} = -\frac{1}{\beta} \ln \langle e^{-\beta \Delta V_{BA}} \rangle_A \tag{2.14}$$

The one-sided perturbation is, however, biased by the exclusive sampling of states in only ensemble A. In order to eliminate this bias, information of both ensembles can be combined with a two-sided perturbation approach where data from both systems is included. The drawback of this approach is the need of generating trajectories for both ensembles A and B, which in general doubles the computational cost. With the two simulations the free energy can then be calculated as

$$\Delta F_{AB} = -\frac{1}{\beta} \ln \frac{\langle e^{-\beta V_B} \rangle_A}{\langle e^{-\beta V_A} \rangle_B} \tag{2.15}$$

In order to decrease the statistical error of the one- and two-sided free energy perturbation methods, Charles H. Bennett came up with a formulation of the free energy difference between two systems as the acceptance ratio of switching the Hamiltonian function from A to B[146]. In contrast to Metropolis who formulated a criterion to describe the acceptance of a set of phase space coordinates in Monte Carlo simulations, Bennett asked for an acceptance criterion which provides minimal error for the calculated free energy difference. For this he expanded the fraction of partition functions which forms the core of difference free energy estimates,

$$\frac{Q_A}{Q_B} = \frac{\langle W(\mathbf{q}) \exp\left[-\beta V_A\right] \rangle_B}{\langle W(\mathbf{q}) \exp\left[-\beta V_B\right] \rangle_A} \tag{2.16}$$

where W can be an arbitrary weighting function. Applying Lagrange multipliers, he found, that the weighting function which provides the smallest errors for a given set of sampling points for both systems is the Fermi Dirac distribu-

tion function $f(x) = 1/(1 + \exp(\beta x))$. With a simple efficiency argument he further showed that the choice for the number of probes in A and B should be equal in case that the computational effort for the creation of one sample equals in both systems. The Bennett acceptance ratio (BAR) for the free energy is thus

$$\Delta F_{AB} = \frac{1}{\beta} \ln \frac{\langle f(V_A - V_B + C) \rangle_A}{\langle f(V_B - V_A - C) \rangle_B} + C. \tag{2.17}$$

The constant C has to be determined in a post-processing step after the simulation with the additional condition

$$\sum_A f(V_A - V_B + C) = \sum_B f(V_B - V_A - C). \tag{2.18}$$

By the time of its invention, the post processing minimization of C was seen as a drawback of the BAR method but has become a negligible effort in recent free energy calculation protocols[147].

In the limit of exhaustive sampling the perturbation method gives the exact free energy change associated with the perturbation. However, the efficiency or convergence of calculated free energies depends on how well the sampled states for the unperturbed system represent relevant states (those with high Boltzmann probability) of the perturbed system. In order to solve this problem in practical applications one often splits the process into a series of N small perturbations and controls the switch from A to B with a coupling coordinate λ. At $\lambda = 0$ the system is represented by a Hamiltonian $\mathcal{H}_A$ representing state A and at $\lambda = 1$ by Hamiltonian $\mathcal{H}_B$ representing state B. At each stage the free energy change associated with a perturbation step is calculated using the force field change up to a given stage as reference. In this way the total free energy change of going from state A to B is given by the sum over all perturbation steps:

$$\Delta F_{AB} = \sum_{i=0}^{N-1} \Delta F_{\lambda_i, \lambda_{i+1}} \tag{2.19}$$

The $\Delta F_{\lambda_i, \lambda_{i+1}}$ represent the free energy changes associated with the transition from intermediate state λ_i to λ_{i+1}. The parameter λ can be coupled in various

ways to the change in the Hamiltonian as discussed in the next section.

Whenever a smooth transition coordinate λ for transforming the Hamiltonian from system A to B is defined, the thermodynamic integration (TI)[148–150] method is applicable which is probably the most common method to extract free energy differences from simulations. It requires the derivative of the Hamiltonian with respect to a control parameter λ associated with the transition for a force field representing state A and state B. The ensemble average of the derivative of the Hamiltonian vs. the control parameter λ corresponds to the derivative of the free energy vs. λ. Often, the analytic λ derivative of the force field can be calculated. By integration of the ensemble average one can calculate the total free energy change for the transition from A to B:

$$\Delta F_{AB} = \int_0^1 \langle \frac{\partial \mathcal{H}(\lambda)}{\partial \lambda} \rangle_\lambda \mathrm{d}\lambda \tag{2.20}$$

The numerical integration can be performed with different methods e.g. trapezoidal, Simpson or Clenshaw–Curtis integration[147] for which the continuous coupling coordinate λ needs to be discretized and the system to be sampled at each of the intermediate Hamiltonians as it is done for FEP in equation 2.19. Several variants of the TI scheme are available. E.g. in the slow growth method the coupling parameter increases with time during a single trajectory, starting from system A and transforming to system B. If the increment of λ is performed too fast, the system may not be in equilibrium and bias from previous λ steps is observed. This is a general problem of the slow growth method and results in a hysteresis when performing the switching in both directions $\lambda(0 \rightarrow 1)$ and $\lambda(1 \rightarrow 0)$. However, as the free energy is a thermodynamic state function, its absolute value should be independent of the direction of the pathway.

In standard TI simulations one tries to equilibrate the simulation system at each discrete λ in order to calculate an ensemble average for the equilibrated system. The influence of different numerical integration methods on the efficiency and accuracy of TI, i.e. Simpson, trapezoidal and Clenshaw–Curtis in comparison with the BAR method, has recently been investigated by Oostenbrink et al. for the serine protease trypsin with four benzamidine-like inhibitors[147]. BAR was found to better handle fewer intermediate steps com-

pared to TI wherein Simpson was found superior to trapezoidal integration. The benefit of non-equidistant spaced intermediate steps as required by the Clenshaw–Curtis integration method depends strongly on the functional form of the integrand $\mathrm{d}\mathcal{H}/\mathrm{d}\lambda$. Overall BAR should be preferred over TI calculations because of its efficiency and robustness with respect to the choice of intermediate steps between A and B[147,151].

Since free energy is a state function the pathway from thermodynamic state A to B is arbitrary. However, in practice the pathway of transforming the Hamiltonian representing state A to B can have significant influence on the convergence and accuracy of calculated free energy differences. As indicated above the most simple method is the linear coupling of the change in Hamiltonian with the control parameter λ:

$$\mathcal{H}(\lambda) = (1-\lambda)\mathcal{H}_A + \lambda\mathcal{H}_B \tag{2.21}$$

In many cases the free energy change is not uniform along the reaction coordinate λ and it is useful to apply a nonlinear coupling:

$$\mathcal{H}(\lambda) = (1-\lambda)^n\mathcal{H}_A(\lambda) + \lambda^n\mathcal{H}_B(\lambda) \tag{2.22}$$

with free choice of the exponent parameter n. The nonlinear scaling is advantageous for scaling potentials with a steep distance dependence such as a repulsive Lennard–Jones contribution. To increase the phase space overlap of states A and B the transition Hamiltonian is typically sampled at intermediate steps of λ. An estimate of the free energy difference of neighboring λ simulations can be derived from either free energy perturbation or thermodynamic integration. Sufficient simulation time has to elapse among the samples of the derivative Hamiltonian $\mathrm{d}\mathcal{H}/\mathrm{d}\lambda$ for TI or the difference Hamiltonian to neighboring windows $\Delta\mathcal{H}$ to provide unbiased samples.

Especially at the end-points of free energy simulations, when potential functions for atoms are created or annihilated, the form of the radial nonbonded interactions results in a singularity of the interaction at close distance between particles. To avoid the associated large forces and large derivatives of the free energy vs. λ the separation-shifted-scaling[152,153] or soft-core scaling

method has become a standard approach. The soft-core interaction function replaces the distance of particles with an effective radius r_A and r_B to remove the singularity in the Lennard Jones and Coulomb interactions for λ values close to 1 or 0, so the soft-core hamiltonian reads as

$$\begin{aligned} \mathcal{H}_{\text{soft-core}}(r) &= (1-\lambda)\mathcal{H}_A(r_A) + \lambda\mathcal{H}_B(r_B) \\ r_A &= \left(\alpha\sigma_A^6\lambda^p + r^6\right)^{\frac{1}{6}} \\ r_B &= \left(\alpha\sigma_B^6(1-\lambda)^p + r^6\right)^{\frac{1}{6}} \end{aligned} \tag{2.23}$$

The soft-core power p, the interaction radius σ and the scaling factor α are parameters to adjust the smoothness of the transition pathway. Recently, alternative formulations for the soft-core scaling scheme have been proposed that alleviate spurious additional minima due to the original formulation[154] or produce a low energy transition pathway[155].

In case of relative free energy calculations of transforming one chemical group into another one can distinguish two types of pathways. In the single topology method atom types representing one chemical group (e.g. thermodynamic state A) are transformed into other types that represent state B. Note that for this pathway it may also be necessary to smoothly transform the bonded geometric description of one group into the bonded topology of the target group. In case of a surplus of atoms in one group it is possible to transform atoms to non-interacting dummies (these dummies still have kinetic energy but all non-bonded interactions are switched off). In contrast, in the dual topology method all atoms representing one group are transformed to dummy atoms and simultaneously atoms of the target group are created (starting from non-interacting dummies, see Figure 2.4 – II). Due to the unphysical transition pathway between A and B this method is often termed alchemical free energy calculation.

2.2.3. Umbrella Sampling Along a Dissociation Pathway

Beside of alchemical transformations it is also possible to obtain the absolute binding free energy of a ligand bound to a receptor molecule from a simulation

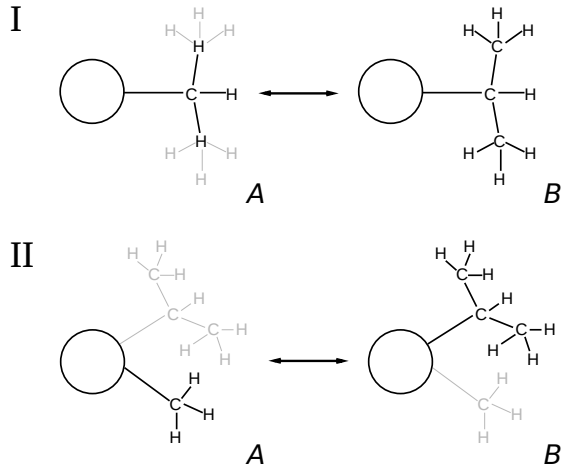


Figure 2.4.: Alchemical transformation of alanine to valine with the single (I) and dual (II) topology approach. With a single topology the atoms extending alanine to valine are gradually turned on. For the dual topology the side chains of valine and alanine are linked to the backbone concurrently. In state A the valine side chain exists as a non interacting dummy and interactions with the rest of the system are turned on while switching the system to state B. Interactions between both side chains are excluded.

of the association or dissociation process itself. From the perspective of the thermodynamic cycle we follow then in principle the natural pathway of studying biomolecular interactions that is the spatial separation of both binding partners, corresponding to the vertical pathways in Figure 2.3.

Instead of a reaction coordinate λ that annihilates or creates the interaction between ligand and receptor a spatial reaction coordinate is chosen to dissociate the ligand–receptor complex. The integration of the generalized mean force along such reaction coordinate yields the potential of mean force (PMF) or work of dissociation. Together with the free energy of release of the ligand into the bulk it allows the calculation of absolute binding free energies. Historically, the PMF was defined as the reversible work supplied to bring two solvated particles from infinite separation to a contact distance:[156]

$$w(d) = -\frac{1}{\beta} \ln(g(d)) \tag{2.24}$$

with $g(d)$ being the pair correlation function of the two particles. The term

"PMF" has since been used for numerous different reaction coordinates, many of which are more intricate than distances. A more general definition[157] of the PMF as used today is the free energy of a state defined by a particular reaction coordinate:

$$W(R) = -\frac{1}{\beta} \ln P_R + W_0 \tag{2.25}$$

where W_0 is a constant and P_R is the probability of the system to be in state R. The free energy difference in direction R can be expressed as the difference of PMF values:

$$\Delta W(R_1 \to R_2) = W(R_2) - W(R_1) \tag{2.26}$$

From a molecular dynamics or Monte-Carlo simulation of a system the free energy difference for a ligand binding to a binding site from the bulk can be calculated as a fraction of Boltzmann weighted states, i.e. probabilities of states:

$$\Delta G_{\text{bind}} = -\frac{1}{\beta} \ln \int_{\text{site}} \mathrm{d}\mathbf{r} e^{-\beta \mathcal{H}(\mathbf{r})} - \frac{1}{\beta} \ln \int_{\text{bulk}} \mathrm{d}\mathbf{r} e^{-\beta \mathcal{H}(\mathbf{r})} \tag{2.27}$$

$$= -\frac{1}{\beta} \ln \left(\frac{\int_{\text{site}} \mathrm{d}\mathbf{r} e^{-\beta \mathcal{H}(\mathbf{r})}}{\int_{\text{bulk}} \mathrm{d}\mathbf{r} e^{-\beta \mathcal{H}(\mathbf{r})}} \right) \tag{2.28}$$

If the ligand is moved solely along the reaction coordinate and the impact of the surroundings along orthogonal coordinates is averaged out, then the free energy of the system, and consequently the probability of states, can be expressed using the PMF:

$$\Delta G_{\text{bind}} = \frac{1}{\beta} \ln \left(\frac{\int_{\text{site}} \mathrm{d}R e^{-\beta W(R)}}{\int_{\text{bulk}} \mathrm{d}R e^{-\beta W(R)}} \right) \tag{2.29}$$

The PMF can be calculated[157–159] as the potential generating the average force acting in the direction R. Forces have to be calculated every step in MD simulations, making them readily accessible, and can be projected onto the direction of R. Commonly the instantaneous forces F_R are collected in

small bins and the PMF can be obtained through numerical integration. This approach is equivalent to thermodynamic integration discussed above for the case of alchemical transformation.

Instead of integration of an average force the PMF can also be obtained directly[160,161] from the probability of states at the coordinate R^*:

$$W(R^*) = -\frac{1}{\beta}\ln(P_{R^*}) + W_0 = W_0 - \frac{1}{\beta}\ln\left(\frac{\int \mathrm{d}\mathbf{r}\delta(R(\mathbf{r}) - R^*)e^{-\beta U(\mathbf{r})}}{\int \mathrm{d}\mathbf{r}e^{-\beta U(\mathbf{r})}}\right) \quad (2.30)$$

Unfortunately, for most systems it is impossible to sample the full conformational space by a single MD simulation on a feasible time scale. Transitions from the bound to the unbound state, if separated by a large energy barrier, will most probably never occur. If the starting point is chosen in the bound state the ligand will not leave the binding site in most simulations and yield no information on the PMF outside of the binding pocket. On the other extreme, a ligand in solution will most probably never enter the binding site. So, simulating all regions of the reaction coordinate for roughly the same time is desirable. To improve sampling along the reaction coordinate the Umbrella Sampling[162,163] method (US) simulates the system with a biased potential. Multiple simulations are run with different potentials u_i. Commonly these potentials are "umbrella-shaped" simple harmonic potentials $u_i = k(R - R_i)^2$, hence the name umbrella sampling (US). Every simulation will restrain the ligand to a small area around the umbrella minimum and will only yield reliable information in that area. Combining these windows then gives a full and effective sampling of the entire association or dissociation pathway along the reaction coordinate. Subtraction of the contribution due to the penalty potential allows the extraction of a free energy within each umbrella window except for an free energy offset between each window. The standard technique to obtain the full free energy change along the coordinate R is the iterative weighted histogram analysis method (WHAM)[164]. Recently, alternative methods have been developed either based on calculating the derivative of the PMF vs. coordinate R in each window[165] or by fitting the sampled data to a free energy curve as a spline function representation[166].

In general, because the conformational space orthogonal to the sampling path R is large, convergence of free energies calculated with US is slow. A solution suggested by Woo & Roux[167] substantially shrinks this space by restraining the ligand to a defined conformation (c), orientation (r) and axis (a) leading from the binding site to the bulk. The conformation is approximately kept fixed by a harmonic potential for the root mean square deviation (RMSD) of the ligand. For orientation and position restraints a coordinate system is introduced based on three centers in the protein and ligand each. The PMF calculation is then performed in umbrella windows pulling the constrained ligand from the bound position into the bulk along the axis. Due to the restraints it is not necessary for each point along the PMF sampling path to sample all possible conformations and orientations of the ligand, which accelerates the PMF convergence. The contributions from the added restraining potentials need to be evaluated at the end points of the PMF calculation, i.e. the ligand being at the binding site and in the bulk. The total free energy of binding is then given by (see Figure 2.5)

$$\Delta G_{\text{bind}} = -\Delta G_{\text{c}}^{\text{site}} - \Delta G_{\text{r}}^{\text{site}} - \Delta G_{\text{a}}^{\text{site}} + \Delta G_{\text{PMF}} + \Delta G_{\text{a}}^{\text{bulk}} + \Delta G_{\text{r}}^{\text{bulk}} + \Delta G_{\text{c}}^{\text{bulk}}. \tag{2.31}$$

The release of the orientational and axial restraining in the bulk can be calculated analytically, whereas other contributions need to be evaluated by numerical FEP free energy simulations. Details on how to best calculate the contributions are given in [167, 168].

The calculation of absolute free energies of binding using the PMF method along a spatial dissociation pathway does not require the annihilation of the ligand in the binding site and in the bulk solvent. The latter two calculations are necessary when choosing the alchemical transformation pathway for absolute binding free energy calculations (see previous paragraph) and depending on the size and properties may result in fairly large calculated free energies. The binding free energy is then obtained as a difference between these two large numbers which may cause significant errors in the calculated free energy. On the down side the PMF approach requires a sterically accessible pathway for dissociation. Otherwise large free energy barriers may also cause errors in the

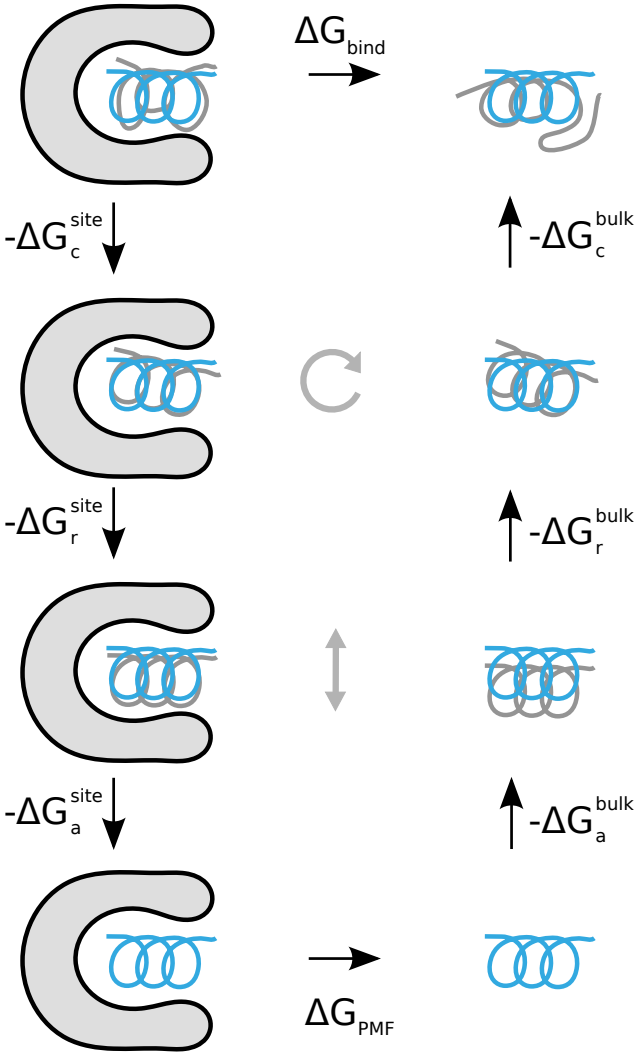


Figure 2.5.: Contributions for calculating the free energy using a PMF approach and restraining conformation (c), rotation (r) and position relative to an axis (a). The left side shows the ligand in the binding pocket, on the right side the ligand is far enough in the bulk that interactions with the protein are negligible. Contributions to the free energy of binding are illustrated only for the ligand. Note, that similar additional contributions also need to be evaluated for the unbound protein.

calculated PMFs. Nevertheless, the PMF approach and several methodological variants have been increasingly applied in recent years[169–174].

2.2.4. Alternative Methods

As indicated above a typical free energy simulation to obtain binding free energy differences due to ligand modifications involves a stepwise transformation with several simulations of intermediate states. It is desirable to reduce the number of intermediate steps in order to reduce the computational demand. Ideally, a transformation should involve only one step. For a one-step transformation (based on one reference state) it would be possible to evaluate several possible modifications simultaneously. Employing the FEP formalism just one simulation of the unperturbed reference state is in principle sufficient to evaluate any modification if the system perturbation is "small" enough to guarantee sufficient overlap with the sampling of states in the reference simulation. The reference state does not need to be an unmodified ligand but can also be an unphysical reference state. The main challenge is here to identify a best possible reference state. Starting in 1996 the van Gunsteren and Oostenbrink groups have put considerable effort into designing efficient schemes for single step perturbation free energy calculations of ligand-receptor systems[175–177]. Typically, soft core centers attached to the unmodified centers are used as reference states since they allow for some overlap of the region, For example, some areas could be accessible for the unmodified ligand byt potentially completely inaccessible for the modified ligand with an added chemical group. Depending on the choice of the reference state and on the type of modification quite accurate estimates of the free energy associated with a chemical modification of a ligand can be achieved[178]. The great advantage is, that it is possible to estimate the effect of various possible modifications (e.g. different chemical groups in the ligand) simultaneously. Each modification represents a perturbation of the chosen reference state. In order to correct for the drawback of a single sided FEP approach to include only information of one system, a method has been devised to combine information of both systems A and B for the free energy difference estimate[178].

Statistical averages are taken for the exponential energy of system B sampled

under the Hamiltonian of A and vice versa. The two sided approach, however, doubles the computational demand because trajectories for both systems have to be generated.

Another class of free energy calculation methods is based on the non-equilibrium work theorem by Jarzynski[179]. In this approach transitions from A to B can be performed rapidly without requiring any equilibration of the system. The associated free energy can be obtained from the average Boltzmann weight of the non-equilibrium work for switching on the transformation. It has also been shown that non-equilibrium approaches to calculate free energy differences are not necessarily more efficient than equilibrium methods[180,181]. A number of variants of the basic methodology have been developed in recent years but the method is still less popular compared to standard TI calculations or alternative equilibrium free energy simulation methods[182].

2.2.5. Current Drawbacks of Free Energy Simulations

In principle, if appropriately performed, the methods discussed in the previous paragraphs allow accurate extraction of free energy differences from molecular simulations. However, as indicated in the beginning of this section, there are several obstacles that limit the applicability and accuracy of free energy simulations. The underlying force field description is one major limitation. For proteins and nucleic acids fairly accurate force fields are available and are evaluated and improved continuously by a large community of users. One problem of current force fields, namely their applicability to intrinsically disordered proteins, will be discussed in Section 2.4. Also, for organic drug-like molecules, the availability of force field parameters is limited. In recent years several approaches for an automatic design and setup of force field parameters for organic molecules have been developed[113,144] The availability of such approaches is a prerequisite for a greater applicability of free energy simulations to evaluate ligand binding and the effect of ligand modification on binding properties. Although most of the common software packages for performing molecular mechanics and molecular dynamics simulations of biomolecules such as Amber[112], Charmm[183], GROMACS[184] or NAMD[185] include the necessary code for free energy calculation, the initiation and setup

can be quite complicated, which may also prevent users from applying free energy simulations[186]. Several efforts to simplify the setup of free energy simulations have been made[187,188].

2.3. Enhanced Sampling Methods

Appropriate sampling of relevant conformational states is a prerequisite for meaningful conclusions to be drawn from MD simulations. Advanced sampling methods can be employed to improve the search for relevant conformations. One of the most common techniques to improve sampling is the replica exchange or parallel tempering method (T-REMD). In the standard setup several parallel simulations, the replicas, are running at slightly different temperatures and frequent exchanges of conformations between simulations at neighboring temperatures are attempted according to the Metropolis criterion

$$\min\left(1, \exp\left\{-\frac{1}{k_{\mathrm{B}}T_1}\left(U(j)-U(i)\right)-\frac{1}{k_{\mathrm{B}}T_2}\left(U(j)-U(i)\right)\right\}\right) \tag{2.32}$$

where $U(x)$ is the potential energy of the coordinates of replica x. An example of exchanges between replicas containing different conformational states is shown in Figure 2.6.

The simulations at higher temperatures can help to overcome energy barriers and, through exchanges, can also help to improve the sampling at lower temperature replicas. Usually, the lowest temperature replica corresponds to the target temperature of interest. Depending on the system size the T-REMD technique can require a significant number of temperature replicas (typically between 8-128 replicas), entailing a significant computational cost for such an approach[190]. However, since all simulations run independently, it is possible to use many parallel simulations on a cluster with little inter-process communication.

A method to reduce the number of necessary replicas compared to T-REMD and still improve the sampling is the Hamiltonian replica exchange method

The contents of this section have been published[189] in similar form

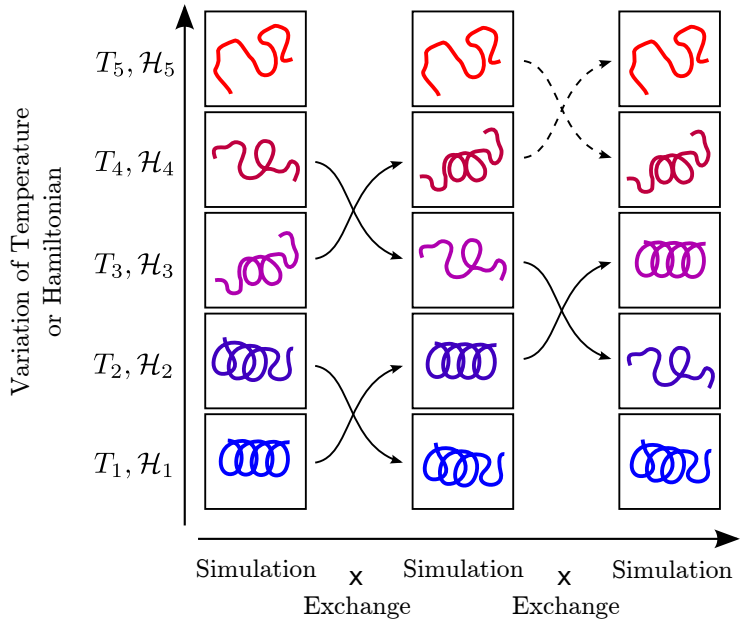


Figure 2.6.: Exchanges between replicas of different temperature or Hamiltonian.

(H-REMD)[191]. Instead of increasing the temperature of the entire system only relevant parts of the Hamiltonian are altered along a control parameter in the replicas. Just as with alchemical transformations the change of the Hamiltonian can be split in several λ-windows to increase phase space overlap and thus exchange probabilities. At frequent intervals conformations in neighboring windows along the control parameter are swapped and exchanges are accepted or rejected according to an adjusted Metropolis criterion:

$$\min\left(1, \exp\left\{-\frac{1}{k_B T}\left(\mathcal{H}_{\lambda_n}(j) - \mathcal{H}_{\lambda_n}(i) - \mathcal{H}_{\lambda_{n+1}}(j) + \mathcal{H}_{\lambda_{n+1}}(i)\right)\right\}\right) \tag{2.33}$$

The difference in the Hamiltonian can, for example, be given by the λ-dependent force field term in alchemical transformations or by the penalty potential associated with the reaction coordinate with US. But any change of Hamiltonian is, in principle, possible. Exchanges of conformational states between US intervals or λ steps can improve sampling at each step and consequently result in better convergence of the trajectories. The quality

of the sampling enhancement is, however, critically dependent on the choice of the control parameter, which can itself restrict the sampling of relevant states. For example, in the case where a US penalty potential is applied to limit the sampling to a certain interval along the reaction coordinate. The penalty potential may constitute a barrier in the system that prevents easy transitions between regimes that are relevant for a given interval along the reaction coordinate. Similarly, in alchemical transformations the sampling at a given λ might be trapped in a certain conformational regime which differs from the regimes sampled in neighboring λ steps.

The H-REMD methodology comes at no additional computational costs and can improve the convergence of free energy simulations considerably[192,193]. It has been used in many areas from PMF-based free energy simulations to dissociate a ligand from a receptor[169,174] and alchemical free energy simulations[191,192] to the folding of small peptides[194] and becomes more and more a standard approach when enhanced sampling is required, especially in free energy calculations. A combination of H-REMD and T-REMD protocols along the one dimensional λ coordinate has been proposed by Wang et al.[195] where the temperature in the transition λ values was gradually increased compared to the end points in order to improve the transition of global configurations.

It is also possible to use λ itself as an additional dynamical variable associated with a mass and kinetic energy[196–198]. In the λ-dynamics approach the reaction coordinate λ is not constant but can vary in between boundaries (usually 0 and 1 for λ) during a single MD simulation[196]. The free energy change along the coordinate can be obtained by integrating the generalized force along the coordinate during the simulation (similar to TI). It is, however, necessary to guarantee sufficient sampling of all relevant λ-values. This may involve appropriate biasing potentials or special sampling techniques[196]. The approach has been applied successfully in a number of examples[197,198].

In the meta-dynamics[199–202] methodology, small Gaussian functions are continuously added along the reaction coordinate to destabilize the currently sampled states. Eventually, this results in a flat free energy surface along the reaction coordinate and the sum of the Gaussian functions represents the free energy curve along the reaction coordinate (with opposite sign). The

convergence to a well–defined free energy surface depends on rapid final diffusion along the reaction coordinate. In order to improve convergence, it has been coupled with other advanced sampling techniques such as T-REMD[169]. The technique has been used in applications from protein folding[203] via protein–protein interactions[204] to ligand binding[204] and the diffusion of ions in channels[205].

For free energy calculations other approaches to improve the sampling of relevant states are based on the inspection of the calculated free energy with respect to the reaction coordinate. A biasing force can be calculated to offset the free energy change (or derivative) along the reaction coordinate. The adaptive biasing force (ABF) reduces the overall free energy gradients along the reaction coordinate resulting in a smoother free energy surface and in more rapid convergence[157,158,206]. In combination with the restraint PMF-methodology described above, the technique has been used to calculate free energies of protein-ligand[167,171] and protein-protein binding[207].

2.3.1. Distance Deviations to Describe Unfolding

Intrinsically disordered proteins, the main subject of this work, can be in a stable conformation for microseconds. Classic continuous MD simulations can access this timescale for the desired system size and thus could capture the transition from one conformation to another. Reliable statistical averages and population probabilities, however, cannot be extracted if transitions between the states are rare. To gain any insight on the statistics of disordered proteins, enhanced sampling techniques have to be used. T-REMD does not lend itself in this case, as the investigated system sizes would require prohibitively many replicas to reach temperatures high enough to speed up unfolding. So, in the method applied here, H-REMD was used.

The crucial choice for any H-REMD method is the selection of a reaction coordinate. In the case of unfolding proteins or peptides, the coordinate should reliably distinguish between the folded state and unfolded states. The RMSD from the folded state, as used in the method of Woo & Roux[167], is a possible choice. It does, however, require a fit to the reference structure for each frame. Instead of taking the RMSD of the coordinates of atoms, here we used the

RMSD of a chosen set of distances compared to respective reference distances, which can be taken from the reference structure (exemplary bonds in Figure 2.7). This dRMSD R is defined as

$$R(d_1, \ldots, d_N) = \sqrt{\frac{1}{N}\sum_{i}^{N}(d_i - d_{i,0})^2} \tag{2.34}$$

Using distances avoids any fitting as the intra-molecular distances are rotation and translation invariant. In addition, the area where conformational freedom is of interest can be freely chosen by the definition of the dRMSD-pairs. With the same mechanism a single helical fragment of a protein can be unfolded or entire domains can be moved with respect to one another.

To enhance sampling along the dRMSD, harmonic potentials force the system to sample specific regions of R around a reference value R_0. The potentials with a force constant k_0 are of the form

$$V(d_1, \ldots, d_N) = \frac{1}{2}k_0 \left(R(d_1, \ldots, d_N) - R_0\right)^2 \tag{2.35}$$

which for pair i between atoms a and b creates a force of

$$\mathbf{F}_a(d_i) = -\frac{k_0}{N} \cdot \frac{R(d_i) - R_{i,0}}{R(d_i)} \cdot (d_i - d_0) \cdot \frac{\mathbf{r}_a - \mathbf{r}_b}{d_i} \tag{2.36}$$

pointing, depending on the sign of the R deviation, towards or away from the bond partner.

Replica exchange between Hamiltonians with different positions of the umbrella minimum further enhances sampling of different conformations and allows the system to overcome artificial barriers introduced by the additional potentials. Note that the phase space volume is not constant with respect to R. While, depending on the number of defined distances, only one or possibly few structures can fulfill $R = 0$, a larger R represents a vast number of conformations.

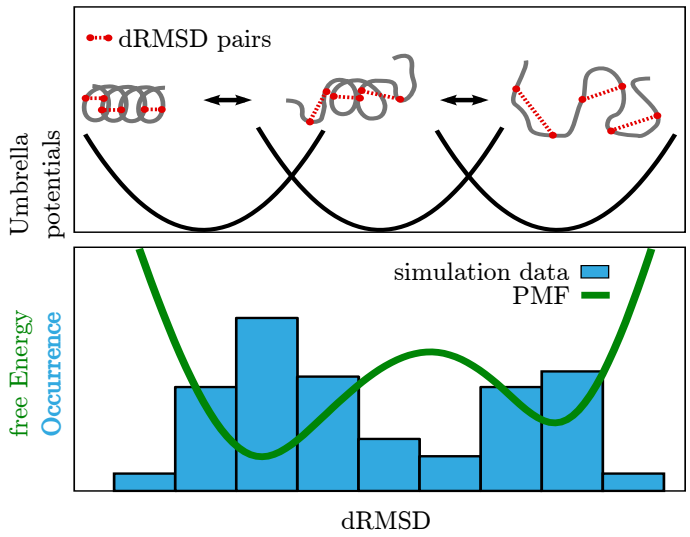


Figure 2.7.: Free energy along the dRMSD reaction coordinate. With harmonic potentials the system is forced to sample all regions of the dRMSD reaction coordinate that describes the average deviation of a chosen set of distances from a respective reference distance. The contribution of the potentials is treated by the WHAM method[164].

2.4. MD Force Fields and IDPs

While computational simulations in principle are an ideal tool for the study of IDPs, the methods still face two major challenges[208]. The first problem is the representative sampling of relevant regions of conformational space. Folded proteins can require long simulation times for convergence, but IDPs with their conformational flexibility and possibly long residence times in certain conformations pose a significantly more difficult task. Studies on ANTON, a computer specifically developed for MD simulations, have shown, that conformational interconversions like the formation and breaking of helical structures of the intrinsically disordered ACBP protein can be on the microsecond timescale[209]. Advances in computer technology like the advent of MD simulations on graphical processing units (GPUs) indicate that increasingly longer time scales are and will be possible. From the picosecond simulations[127] of 1977 to millisecond simulations[210] in 2011 less than 40 years passed for simulation times to increase by 9 orders of magnitude. Extrapolating this number makes it probable that sampling problems for systems that today are

impossible to cover will not be a problem in the future.

The second major challenge is the accuracy of current force field models. Updates have included backbone parameters[211] side chain torsional potentials[212], but were mostly parametrized with and thus optimized for folded proteins. Systematic tests show that newer force fields do perform better than previous versions which they are based on[213,214]. For folded proteins, the amber99SB*-ILDN and charmm22* force fields performed best and reproduced experimental data[213]. When it comes to the dynamic process of protein folding, however, force fields already differ at being able to fold proteins[215], in particular if different classes like α-helical or β-sheet like structures are investigated[213]. Protein–protein interactions might generally be overestimated as simulations with several force fields showed an aggregation of villin headpiece that does not occur in experiment[215]. For IDPs, current force fields have been applied by a number of studies, where in some cases experimental findings were reproduced[216–221], but also questions arose if IDPs and their unfolded states are modelled correctly[209,221–229]. Indeed, in a recent study that compared the generated conformational ensembles, a change of force field had more impact on the generated ensemble than a change of the entire protein sequence[208].

Three major observations point towards insufficient solvation as a possible source of error: Solvation free energies of side chains systematically are too low[230], the association process of proteins is too favorable[231] and dimensions of proteins in explicit solvent are too small compared to experiment[222,228,232]. Water models have been optimized with respect to pure water properties[130,233] and it is little surprising that there remains potential for optimizations. To ameliorate the problem of insufficient sampling of expanded states in IDPs, two force field updates have recently been published.

For Best et al.[228] currently available water models like TIP4P-EW[233] and TIP4P/2005[130] reproduce liquid water properties well. The authors thus conclude that the problem of current force fields to represent IDPs lies in the solute–solvent interactions. In their attempt to alter the carefully balanced current force fields as little as possible they confined adjustments to the Lennard-Jones parameters of the solvent–solute interactions, solely increasing ϵ by a factor γ. A value of $\gamma = 1.1$ was determined as the near optimal factor in

comparisons of the temperature dependent collapse of the CspM34 protein with experimental data from FRET measurements. The resulting force field named amber03ws was tested and compared with experiment for numerous systems. It reproduced SAXS data for an IDP, protein self-association rates, the intrinsic structure propensity in short peptides, helix-coil transitions, the folding of mini-proteins, ameliorated solvation free energies of amino acid analogues and importantly also did not unfold proteins that are naturally folded.

Piana et al.[222] simulated the intrinsically disordered N-terminal zinc binding domain of HIV-1 integrase (IN) with several force fields and water models for between 5 - 141 µs. For all force fields, the resulting radius of gyration was underestimated by a factor of two compared to FRET measurements. Switching from TIP3P to TIP4P-EW or TIP4P/2005 did not show significant improvements. The authors suspect that dispersion components, i.e. the attractive van-der-Waals energy contributions, are systematically underestimated by current water models. While an error in the dispersion could still reproduce the mainly dipole-dominated water-water interactions and thus pure water properties, it might have an effect for water as a solvent and on the behavior of solvated molecules like IDPs. To better estimate dispersion interactions, force fields were compared to quantum level calculations of a set of small dimers representing typical functional groups. All current force fields compared to the quantum data did underrepresent solvent dispersion effects, so as a solution the authors suggest the TIP4P-D force field with the dispersion term of the Lennard-Jones potential, C_6, increased by 50% compared to TIP3P. the remaining water parameters were then fitted to match the density and vaporization enthalpy temperature profiles of water. The water model was tested for a total of 830 µs by checking the solvation free energies of side chains, radii of gyration of disordered proteins and a detailed comparison with SAXS, FRET and NMR PRE data for α-Synuclein.

The two force field corrections have been suggested independently and tackle the problem of under-solvated IDPs in a slightly different way. While solvation free energies of small molecules are left unchanged for TIP4P-D, they are a major motivation for the development of amber03ws. In contrast, properties

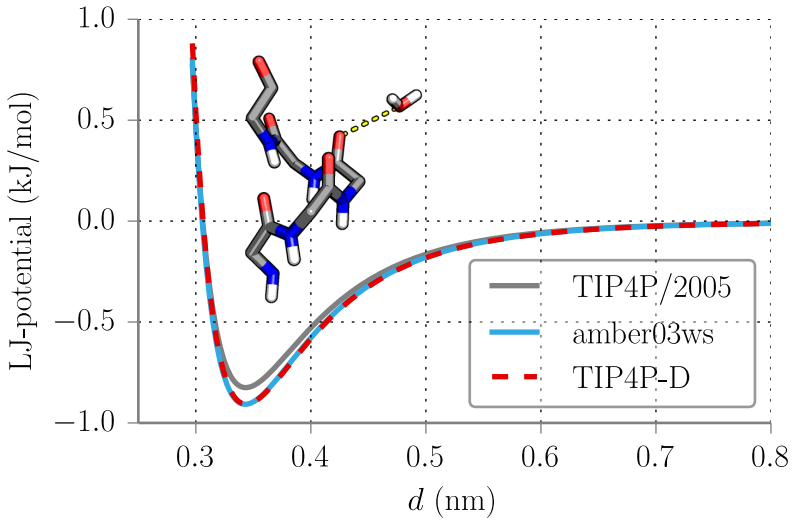


Figure 2.8.: Lennard-Jones potential between a water oxygen and a backbone oxygen with three different water models.

of pure water are altered with TIP4P-D and changed for amber03ws. The net change to water–protein interactions is, however, remarkably similar, as shown for an exemplary potential between the oxygen of a water molecule and a backbone oxygen in Figure 2.8. A more detailed comparison especially of the folding of short peptide sequences will be the topic of chapter 3.

As a final note on the force field issue it is possible that current force field models are in general not accurate enough for principle reasons because important physical effects are not covered. For example, electronic polarizability of the electron clouds is not explicitly included in current force fields. Several new force fields have been developed in recent years to include electronic polarisability at least approximately[234–237]. Their performance with intrinsically disordered proteins has yet to be tested, but the additional computational cost will further impede the sampling problem and make equilibrated conformational ensembles harder to reach.

3. Performance of Force Fields Developed for IDPs

Many natural proteins are as a whole or in part intrinsically disordered. Frequently, these intrinsically disordered regions (IDRs) undergo a transition to a defined and often helical conformation upon binding to partner molecules. The intrinsic propensity of an IDR sequence to fold into a helical conformation already in the absence of a binding partner can have a decisive influence on the binding process and affinity. Using a combination of NMR spectroscopy and molecular dynamics (MD) simulations we have investigated the tendency of regions of Axin-1, an intrinsically disordered scaffolding protein of the WNT signaling pathway, to form helices in segments interacting with binding partners. Secondary chemical shifts from NMR measurements show an increased helical population in these regions. Systematic application of MD advanced sampling approaches on peptide segments of Axin-1 reproduces the experimentally observed tendency and allows insights into the distribution of segment conformations and free energies of helix formation. The results, however, were found to dependent on the force field water model. Recent water models specifically designed for IDRs significantly reduce the predicted helical content and do not improve the agreement with experiment.

3.1. Introduction

Several experimental techniques can be used to qualitatively characterize the disordered nature of proteins[13]. Obtaining information on the ensemble of disordered states or transiently formed structures is, however, experimentally difficult, as typically only averages over time or large sample sizes can be

evaluated. Also IDP systems tend to be underdetermined, i.e. the number of experimental observables is below the number of conformational states. Computational studies are, in principle, ideally suited for the study of IDPs, allowing single molecule investigation with a spatial resolution up to single atoms and a time resolution of femtoseconds.[38] Especially in combination with experimental techniques like nuclear magnetic resonance (NMR) spectroscopy an atomistic understanding of the range of accessible conformations is available[35,36].

As described in chapter 2, molecular dynamics simulations model protein behavior by propagating all atom positions according to Newton's equations of motion[238] and estimating forces on atoms via an empirical force field. Current force fields describe folded globular proteins well[213], but overstabilize protein–protein interactions[231] and often fail to reproduce realistic IDP behavior[208,222,228]. Most force fields succeed in predicting qualitatively whether or not a region is disordered, but for further details, like dynamics or sampled sub-states, results vastly differ between different force fields[208]. Two groups have approached this problem by parameterizing new water models to achieve better. Both approaches tackle the overpopulation of collapsed states with current force fields by modifying the solute–solvent interactions.

Best et al.[228] argue that pure water properties are reproduced well by current water models and thus water–water interactions should not be altered. For their adjusted water, termed amber03ws, the strength of close range solvent–solute interactions is increased by a factor of $\gamma = 1.1$, obtained from fits to the temperature profile of Förster resonance energy transfer (FRET) data for the Csp M34 protein. Extensive tests of the new force field included comparison with Small-Angle X-ray Scattering (SAXS) data, solvation free energies of amino acid analogues, protein self-association, the intrinsic structure propensity in short peptides, the helix-coil transition, the folding of mini-proteins and the stability of folded proteins.

According to a second approach by Piana et al.[222], the dispersion component of the intermolecular interaction energy is underestimated in current water models. They fitted the C_6 term of Lennard-Jones interactions to quantum level computations and adjusted partial charges and C_{12} to fit density and

vaporization enthalpy temperature profiles of their TIP4P-D water model. The water model was tested in a total simulation time of 830 µs by checking the solvation free energies of side chains, radii of gyration of disordered proteins and a comparison with SAXS, FRET and NMR paramagnetic relaxation enhancement data for α-Synuclein. Both water models performed well in test simulations and in particular sampled larger radii of gyration in agreement with experiments.

Aside from radii of gyration MD force fields have to reproduce very diverse properties of disordered regions. Often, disordered regions in proteins are involved in the interaction with other biological binding partners and can fold upon binding adopting different conformations. Among these, helical structure is the most abundant secondary structure. Hence, in practice one is interested in the ability of force field simulations to distinguish segments from IDPs that have an intrinsic preference for adopting helical secondary structure and can bind involving a conformational selection mechanism or prefer fully disordered states. Here we use the intrinsically disordered, transiently folded Axin-1 protein as model system to compare computational approaches for conformational characterization of IDPs.

Axin-1 is a key protein component of the Wnt signaling pathway[239] acting as a scaffolding protein, assembling the b-catenin destruction complex that phosphorylates and subsequently polyubiquitinates β-catenin[240–242]. The central region of the 862 residues Axin-1 protein (residues 212-780) is highly susceptible to proteolytic degradation[243] and has been proposed to be largely intrinsically disordered[244]. Nevertheless, the disordered region is essential for β-catenin trapping and subsequent degradation in the β–catenin destruction complex. For these purposes, the disordered region of Axin-1 harbors binding sites for β-catenin (residues 466-480) and the kinases Glycogen synthase kinase 3β (GSK3β; residues 383-400), and Casein Kinase 1 (CK1)[245–247]. In complex with each of the binding partners, Axin-1 adopts a helical structure in the respective binding region[245,247]. The intermediate region connecting the binding sites (residues 430-450) is proline-rich and predicted to be in a coiled coil state[248].

In this chapter we investigate the helix propensity of the Axin-1 residues

380-490 by segmenting the amino acid sequence in peptides of 10 residues and assessing the conformational space of these peptides with an advanced sampling MD simulation method. Simulation results are directly compared to results obtained from NMR spectroscopy on the Axin-1 segment. Good qualitative agreement between MD simulations and experiment is found indicating that segments that bind to signal proteins indeed adopt already partially helical conformations in absence of the binding partner. However, significant differences between force field description and type of water model used in the simulations are found. While the amber99SBws[228] force field with corrections to the backbone parameters reproduces enhanced helicity in certain sequences, TIP4P-D[222] water does not reproduce transient helical population.

3.2. Methods

3.2.1. Simulation of the Helical Propensity

The relevant section of Axin-1 (residues 380-490) is, at least for current computer power, too large to be simulated for timescales on which the conformational space could be assessed. In order to obtain an estimate of helicity in different areas of Axin, we separated the amino acid sequence into segments of 10 residues. Sequences of these segments are given in Table 3.1. For each of these peptides we ran an advanced sampling MD simulation protocol to estimate the free energy (or potential of mean force, PMF) along a reaction coordinate that maps the helicity of the peptide. As reaction coordinate we used the root mean square deviation of a set of i distances from a respective set of i reference values:

$$R(d_i) = \sqrt{\sum_i (d_i - d_{i,0})^2} \tag{3.1}$$

All distances between C_α atoms $C_{\alpha,j} - C_{\alpha,j+3}$ were included, with a reference $d_{i,0} = 0.5\,\mathrm{nm}$ for every i, meaning that $R = 0$ represents the fully helical peptide and an increase of the R shows increasing deviation from the helical structure.

first residue	sequence
381	VRVEPQKFAE
391	ELIHRLEAVQ
401	RTREAEEKLE
411	ERLKRVRMEE
421	EGEDGDPSSG
431	PPGPCHKLPP
441	APAWHHFPPR
451	CVDMGCAGLR
461	DAHEENPESI
471	LDEHVQRVLR
481	TPGRQSPGPG

Table 3.1.: Sequences of simulated segments

Force Fields

We used the amber99sb*-ILDN force field for the peptides, which is an revised version of amber99sb[249] with improved side chain torsion parameters[212] optimized for helix-coil transitions[250]. For the solvent the classical TIP3P[129] water was compared to new force fields explicitly developed for intrinsically disordered proteins. TIP4P-ws[228] comes with an increased k_Ψ for the protein backbone from 0.75 kJ/mol to 2.0 kJ/mol, as indicated in the respective SI[228]. To compare solely the performance of the water force field we also tested a variant we termed TIP4P-s that does not edit the protein force field by adjusting k_Ψ. The set of water force fields was completed with TIP4P-D[222].

Simulation Protocol

All simulations were conducted using GROMACS 4.6.5[251], applying periodic boundary conditions and covering long-range electrostatic interactions with the Particle-Mesh-Ewald[252] method with a Fourier-spacing of 0.16 nm and a grid interpolation up to order 4. Close Van-der-Waals and Coulomb interactions were cut off at a radius of 1.0 nm. Long range dispersion correction was applied to account for errors from truncated Lennard-Jones interactions. Bond lengths of H-atoms were constrained with the LINCS[253] algorithm and a coupling matrix extension order of 4 (12 in equilibration runs). All systems

were run with a step size of 2 fs at a temperature of 300 K, controlled by velocity rescaling[254], and a pressure of 1.01 bar with the Parrinello-Rahman barostat[255].

Starting structures of the peptides were generated with PyMol[256] in a helical conformation and with ACE and NH2 caps at the ends. The peptides were solvated and in a dodecahedral box large enough to accommodate the fully unfolded molecules and containing, depending on the specific sequence, approximately 2000 water molecules. Initial energy minimization with the steepest descent algorithm was stopped when the maximum force dropped below 100.0 kJ/mol/nm or after 25000 steps. Subsequent equilibration was performed per replica with a time step of 1 fs and the velocity rescaling thermostat[254] for 50.000 steps as NVT equilibration and for 100.000 steps with the Berendsen barostat[120] as NPT equilibration.

For each segment the free energy landscape was sampled in 12 equally spaced λ-windows for values of R between 0.0 - 0.5 nm. Replica exchange between windows was attempted every 500 steps to further enhance convergence and overcome potential artificial energy barriers.

3.2.2. Experimental Procedure

Protein Preparation

The DNA construct containing the Glu390-Val500 region of human Axin-1 (uniprot reference O15169) was purchased from ATG:biosynthetics GmbH within a pUC cloning vector. The DNA sequence was codon optimized for protein production in bacterial cells and flanked by NcoI and BamHI restriction sites. The coding region was cloned into a modified pETM-11 bacterial expression vector (kindly provided by Arie Geerlof Protein Expression and Purification Facility, Helmholtz Zentrum München, Germany) which was derived from a pET-24d(+) vector (Novagen) by insertion of a tobacco etch virus (TEV) protease cleavage site following a N-terminal hexa-histidine and a protein A tag. Axin-1 gene was amplified by PCR using T4 primers (New England Biolabs). The resulting PCR products and pETM-11 were double digested with NcoI and BamHI enzymes (New England Biolabs) before ligation. The construct was verified by sequencing. The numbering of Axin-1

amino acid residues follows the full length protein as reported on the UniProt website. Uniformly (^{13}C,^{15}N) double-labelled protein was produced in freshly transformed *E.coli* DE3 cells. A single colony was inoculated in Luria-Bertani medium (20 ml) with kanamycin (25 mg/l) and cultured at 37 °C until the OD_{600} reached a value between 2 and 3. From this, an aliquot (1ml) was added to (^{13}C,^{15}N labelled) M9 minimal medium (100 ml) in which ^{15}N-NH_4Cl (1 g/l) and ^{13}C-glucose (2 g/l) were the only sources of nitrogen and carbon for NMR isotope labelling purposes (Cambridge Isotope Laboratories, Inc). The culture was incubated overnight at 37 °C and shaken at 180rpm. Fresh (^{13}C,^{15}N) M9 minimal medium was added up to 1 l, and the culture was grown under the same conditions until the OD_{600} reached 0.8. Protein expression was induced with 1 mM β-D-1-thiogalactopyranoside (IPTG) at 18 °C. The cells were pelleted next day by centrifugation using a Fiberlite F9-6x1000 rotor in a Sorvall LYNX 6000 Superspeed centrifuge at 2000 g for 20 minutes.Resuspension and contemporaneous cell lysis were obtained adding 40 mL of 50 mM NaPi, 300 mM NaCl, 8 M urea, pH 8 and gentle stirring at room temperature for 20 minutes. The cell lysate was separated by ultracentrifugation using a Thermo Scientific SS-34 rotor in a Sigma 6K 15 centrifuge at 20 000 g for 30 min at 4 °C and histidine-tagged protein was affinity-purified via Ni-NTA resin (Qiagen). TEV protease (5 µg/ml) was added to the eluate and dialysed overnight at 4 °C against 50 mM Tris(hydroxymethyl)amminomethane chloride (TRIS-HCl), 150 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), 2 mM β-mercaptoethanol, pH 7.5. The dialysed solution underwent heat shocking at 90 °C for 30 minutes, followed by ultracentrifugation using the Thermo Scientific SS-34 rotor in Sigma 6K 15 centrifuge at 20 000 g for 30 min at 4 °C. Axin-1 was separated from the tag by a size-exclusion chromatography step, using a Superdex 75 10/300 GL column (GE Healthcare), equilibrated with 20 mM NaPi, 300 mM NaCl, 2 mM 1,4 Dithiothreitol (DTT), pH 6.5, on an ÄKTA pure FLPC system. The concentration of the protein was estimated by absorption spectroscopy ($\epsilon_{280} = 5.6\,\mathrm{mM}^{-1}\,\mathrm{cm}^{-1}$). The purity was estimated by SDS-PAGE to be 95% with a yield of pure protein being 10 mg per litre of culture.

NMR spectroscopy

Axin-1 sample (0.5 mM) with 0% and 40% TFE contained 20 mM NaPi, 300 mM NaCl, 2 mM DTT, pH 6.5 and 10% D_2O for lock. Triple resonance backbone assignment experiments included: CBCA(CO)NH, CBCANH, HNCA, HNCAN(N)H, HNCANN(H), HCCC(CO)NH and HCCH-TOCSY. All NMR spectra were recorded at 298 K on Avance III 600 MHz and Avance II 900 MHz Bruker spectrometers, equipped with TCI (^{1}H, ^{13}C, ^{15}N) with z-gradient and TXI (^{1}H, ^{13}C, ^{15}N) with z- and xyz-gradient cryoprobes, respectively. Spectra were processed with NMRPipe[257] and analyzed with CcpNMR Analysis[258]. The secondary structure propensity (SSP) was derived by the online tool ncSPC[259] (neighbour-corrected Structural Propensity Calculator), according to

$$\mathrm{SSP} = \Delta\delta C_\alpha - \Delta\delta C_\beta \tag{3.2}$$

where $\Delta\delta C_\alpha$ and $\Delta\delta C_\beta$ represent the difference between carbon α and carbon β Axin-1 experimentally obtained chemical shifts and carbon α and carbon β random-coil chemical shift references, respectively.

3.3. Results

3.3.1. Free Energy Landscape Along the Amino Acid Sequence

The free energy landscape of helix formation for the 11 segments of Axin with four tested water models is shown in Figure 3.1. With TIP3P water the regions of R with the lowest free energy fluctuate from segment to segment. While some segments, residues 390-420, 440-450 and 460-480, have their global minimum close to the helical structure at $R < 0.2$ nm, the other segments prefer unfolded structures with the global minimum around $R \sim 0.3$ nm and above. With TIP4P-D almost all segments show global minima at very high $R \sim 0.4 - 0.5$ nm where the peptide chain is almost fully extended. Shifts of the global free energy minimum in the R-axis between segments are reduced and helical states with small R-values are generally disfavored. The free energy landscape with TIP4P-s favors or disfavors helical structures for the similar

ranges of residues as TIP3P. TIP4P-ws with a backbone correction towards more helical states indeed shifts the minima towards lower values of R.

3.3.2. Population of a Helical State

For *in silico* studies, the definition of a helical state of a peptide is not straightforward. At atomic resolution helicity is often classified using the DSSP algorithm[260], based on an energy function for the backbone hydrogen bond stabilizing the helix. In Figure 3.2 the average DSSP helicity of all residues of all peptide sequences is plotted with respect to R. Overall helicity decreases linearly with R. For $R = 0.15\,\text{nm}$ the curves drop below a total helicity of 50%. This R was subsequently used as an upper boundary for the helical state of a peptide.

however, be in helical dssp state since their respective helix-H-bond partner does not exist. So complete helicity of the 10 residue segment means 0.7 DSSP helicity.

Population of this helical state is displayed in Figure 3.3 for each segment. TIP3P predicts transiently helical segments between residue numbers 390 and 420, a small helical population between residues 440 and 450 and further transient helicity between residues 460 and 480. With TIP4P-s the same residue regions show a relevant population of the helical state, but the 390-420 region is less helical while in the 460-480 region especially the segment starting at residue 471 is more helical than with TIP3P. Simulations with TIP4P-ws reproduce and amplify the helicity peaks of those with TIP4P-s. Finally, with TIP4P-D almost no increase of helical populations in the respective regions can be seen. Note that the position of the helical boundary at $R = 0.15\,\text{nm}$(see Figure 3.2) affects the absolute values of the population, but the relations between segments and force fields persist.

quarter of the simulation.

F-P-A interaction.

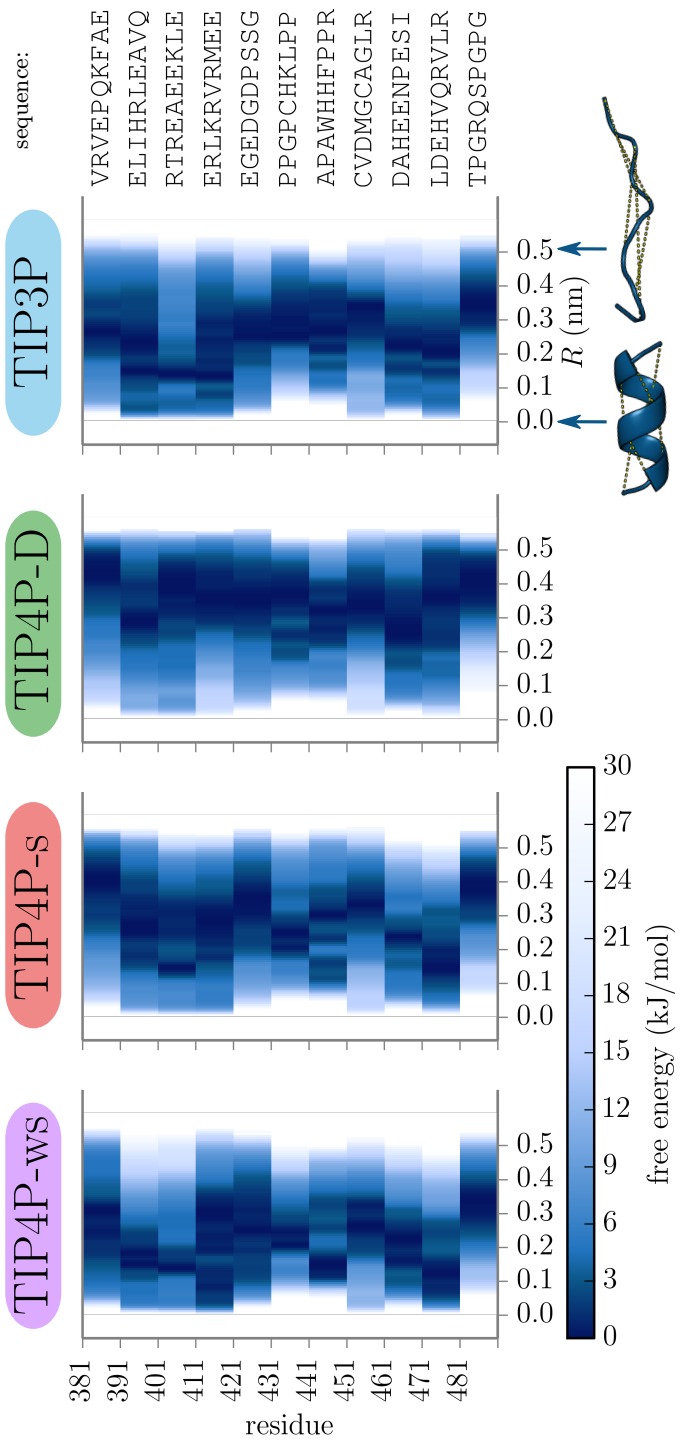


Figure 3.1.: Free energy landscape for segments of Axin along the unfolding reaction coordinate R with different water force fields.

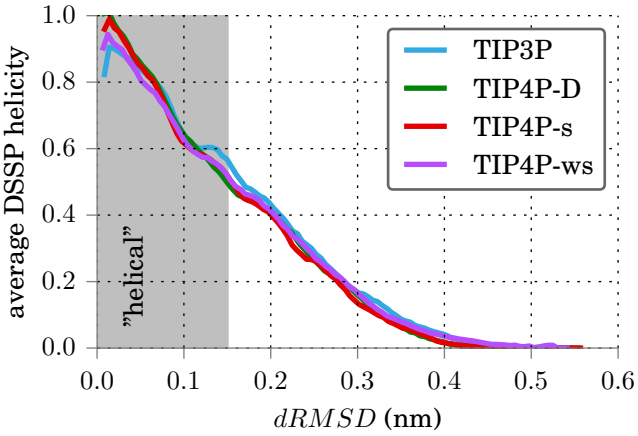


Figure 3.2.: DSSP Helicity of peptides with increasing R for different water force fields. The average is taken over all segments and all replicas for each force field.

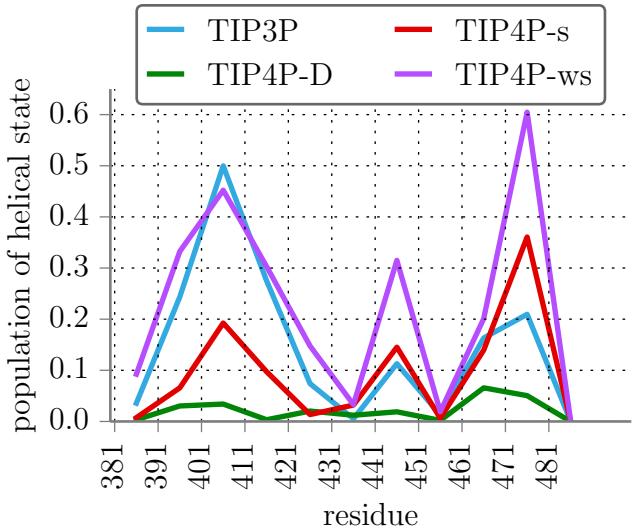


Figure 3.3.: Fraction of population of a helical state for a range of segments of Axin-1 with different water force fields.

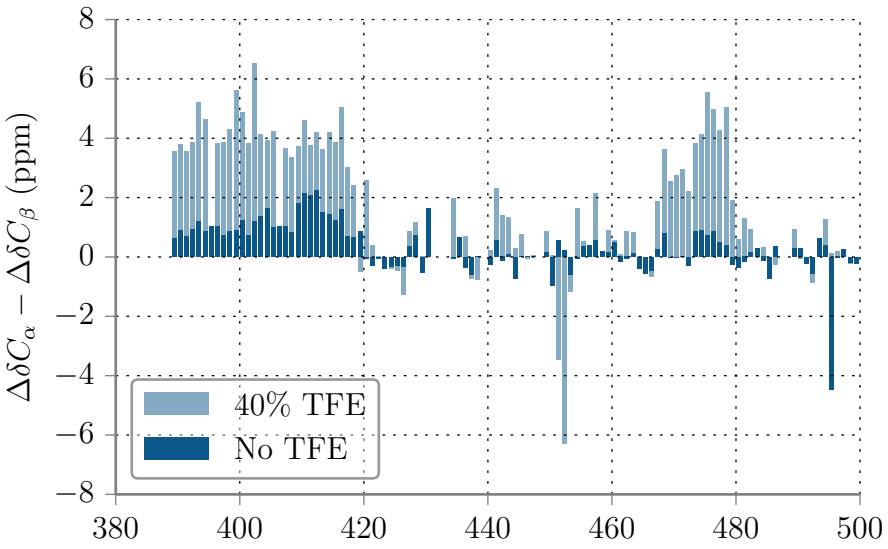


Figure 3.4.: Difference between experimental NMR chemical shifts and calculated chemical shifts of C_α and C_β per residue of Axin. Consecutive positive values indicate α-helices.

3.3.3. Comparison with Experiment

To evaluate the population of secondary structure experimentally, NMR spectroscopy was used. In line with previous work[244], two-dimensional ^{1}H–^{15}N heteronuclear single-quantum coherence (HSQC) NMR correlation spectra recorded on isotope-labeled Axin-1 protein showed that the protein is indeed largely disordered as indicated by clustering of resonances in a narrow range of 7.6 - 8.6 ppm (Figure 3.4). Nevertheless, NMR-derived secondary chemical shifts indicate that both the GSK3β and β-catenin binding sites adopt transient α-helical conformation. In line with this, addition of TFE, an agent stabilizing α-helical structure[261], increased the population of α-helical conformation (Figure 3.4). Based on the secondary NMR chemical shifts obtained in presence of TFE, Axin-1 adopts a helicity of approx. 40% for residues 390-420 and 15% for residues 470-480 in the native state, respectively.

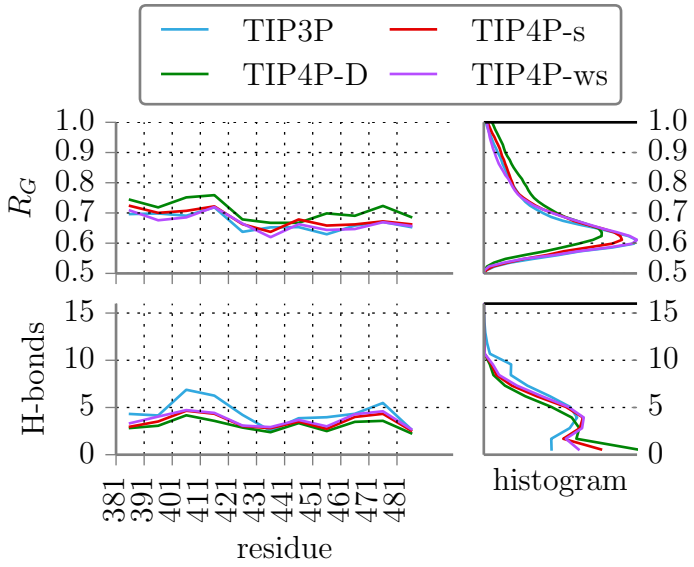


Figure 3.5.: Averages per segment and total histograms of radius of gyration and number of intra-molecular H-bonds of the peptides for different force fields.

3.3.4. Differences Between Force Fields

The simulation results on the peptides using the four water models also indicated differences in the sampled peptide structural properties. Figure 3.5 shows the averages per segment and global histograms of the radius of gyration and number of intra-molecular H-bonds of the peptides in different force fields. In the simulations with TIP4P-D states with a larger radius of gyration were consistently sampled more frequently than with the other water force fields. With TIP3P, instead, more states with a higher number of intra-molecular hydrogen bonds were sampled. Especially for the segments 401-410 and 411-420 many intra-molecular H-bonds persisted. In the segment from residue 401 there are two Arg-Glu pairs that stabilize the helical fold: Arg403-Glu410 and Arg401-Glu404 on opposite sides of the helix (see Figure 3.6). With TIP3P water these two bonds are more persistent than with TIP4P-ws and far more persistent than with TIP4P-D. Similarly, a stable bond between Arg412-Glu420 stabilizes a slightly kinked helical motif and is sampled the least by TIP4P-D.

To find further differences in the sampled structures we clustered the trajecto-

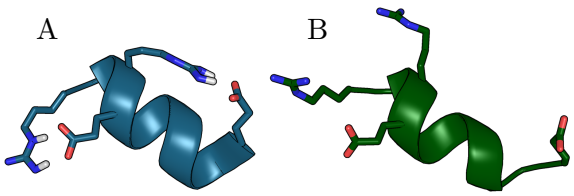


Figure 3.6.: Snapshots from the lowest R replica of segment 401. A shows two stable double-H-bonds between Arg401-Glu404 and Arg403-Glu410 typically sampled with TIP3P water. B shows a typical snapshot from TIP4P-D water where neither contact is formed.

ries of all replicas and force fields. This can provide information on the degree of diversity of sampled structures and can tell whether or not some structures are strongly favored only for a specific force field. Figure 3.7 shows average cluster sizes after sorting the clusters by size for each force field. A tendency of TIP4P-D can be observed to sample highly populated clusters less and sparsely populated clusters more. In a second plot Figure 3.7 furthermore shows the number of populated clusters for each force field and segment. All tested water force fields sampled similar cluster size distributions. The numbers of clusters in the different segments approximately agree for all force fields. Only for the segment starting at residue 391 simulations with TIP3P sample a larger number of clusters than the modified IDP water force fields. Simulations with both TIP4P-ws and TIP4P-s systematically sampled fewer clusters than the other two.

A more detailed look at the obtained clusters involved checking for structures predominantly sampled by specific force fields. We identified all significant clusters (i.e. containing more than 100 structures) of all segments that were dominated or neglected by a force field. Clusters were considered neglected if less than 1% and dominated if more than 80% of the cluster structures was contributed from simulations with one force field. We found 118 dominated clusters (30 TIP3P, 35 TIP4P-D, 27 TIP4P-s, 26 TIP4P-ws) and 378 neglected clusters (68 TIP3P, 93 TIP4P-D, 108 TIP4P-s, 109 TIP4P-ws). In Figure 3.8 the radius of gyration and the number of intra-molecular H-bonds of all such dominated or neglected clusters are plotted. TIP3P dominates for structures with small R_G but with many H-bonds and neglects structures with larger R_G.

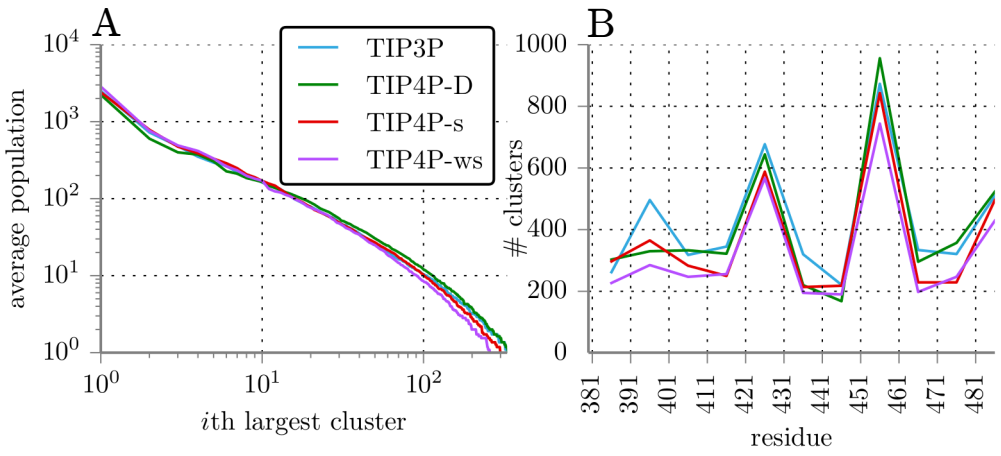


Figure 3.7.: Left: Average size of clusters after sorting clusters by size for each force field. Right: Number of populated clusters for each segment and force field.

TIP4P-D dominates structures with large radii of gyration and neglects several clusters with many H-bonds. TIP4P-s and TIP4P-ws both dominate fewer clusters than the first two water models and neglect more clusters. Dominated clusters are all in the small R_G and modest number of H-bonds regime, and neglected clusters include extended peptides, more than TIP3P, as well as structures with many H-bonds.

3.3.5. Convergence

Figure 3.9 shows the convergence of the PMFs after 80 ns, with the first 10 ns dropped as equilibration time. After the first quarter most PMFs do not shift any more. Some segments, notably starting at residue 431 and 441, still show a continued decrease of the PMF in the unfolded region, but the overall shape of the free energy remains similar. Obtained PMFs from the four different force fields show very similar shapes for each segment except for the emphasis on the helical state and thus are considered converged. The similar numbers of sampled clusters are a further indication that (at least a comparable level of) convergence has been reached in all simulations.

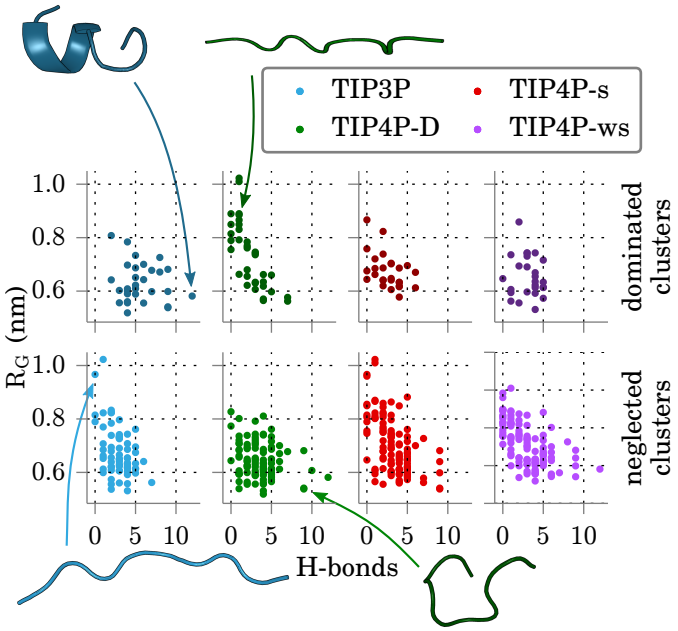


Figure 3.8.: Radii of gyration and number of intra-molecular H-bonds for all clusters dominated or neglected by the different water force fields. Clusters were considered dominated by a force field when more than 80% and neglected if less than 1% of their structures was obtained with that force field. Representatives of indicated clusters are shown as cartoon.

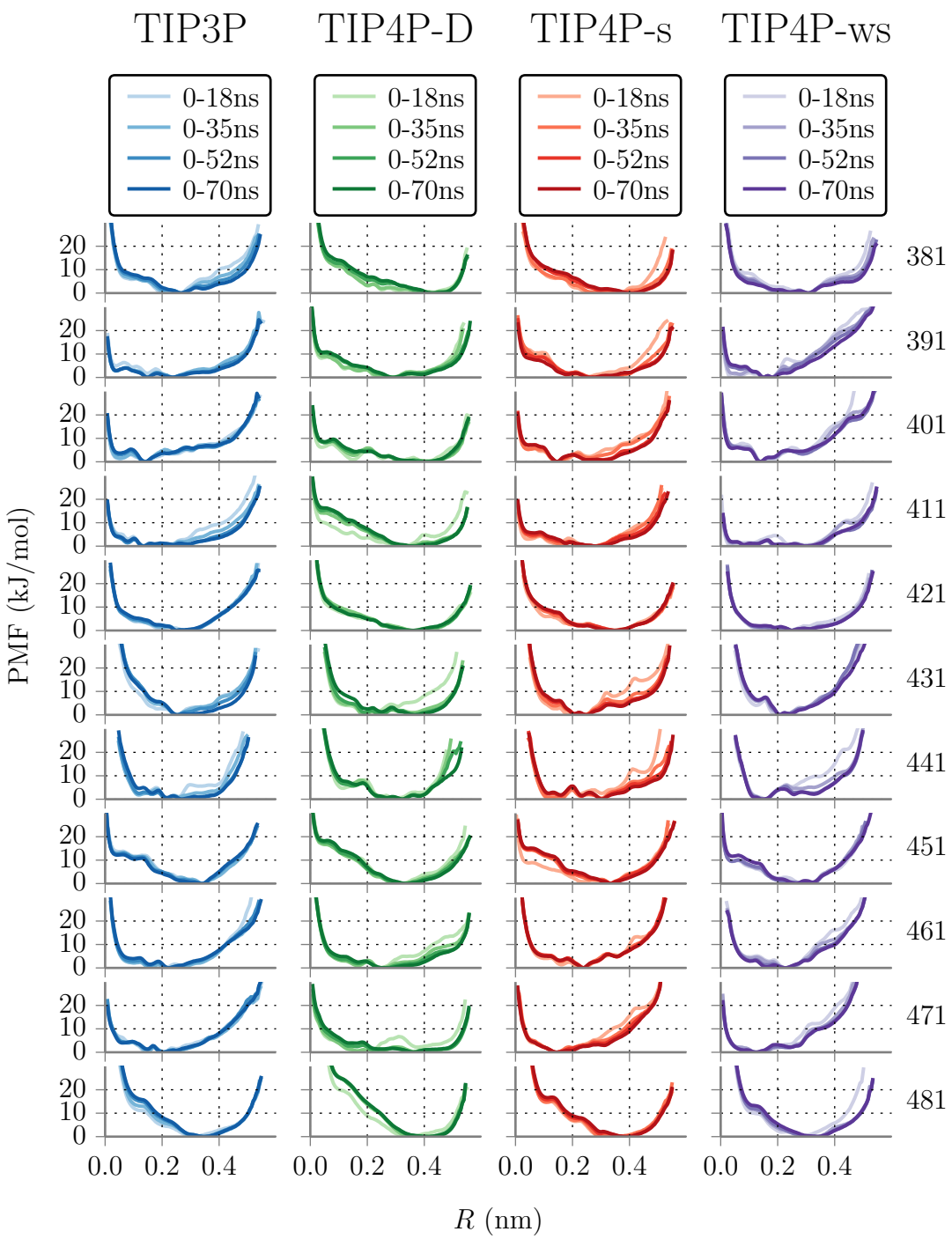


Figure 3.9.: Time evolution of all PMFs for all water force fields. Simulation data was evaluated after 10ns of equilibration. Global shapes of the PMFs barely change after the first quarter of evaluated simulation time.

3.4. Discussion

NMR secondary chemical shifts and MD simulations provided evidence that Axin-1 is largely disordered but shows areas of helix propensity, especially in the binding regions of GSK-3β and β-catenin. The transiently folded regions identified here agree with results from the neural network predictor PONDR[248], where residues $\sim 380 - 400$ and $\sim 450 - 480$ show a reduced disorder score. Segmenting the 100 amino acid chain into peptides of 10 residues we were also able to identify regions of increased helicity in MD simulations. The used water model, however, has a significant impact on the structural ensemble of 10 amino acid peptides.

The TIP3P water model reproduces the higher helix propensity for the binding regions as observed in the NMR experiments. The helix propensity is in the same order as the experimentally observed helix formation probability. Newer water force field models, specifically adapted to describe IDPs, reduce the helicity of all segments. TIP4P-s reduces overall helicities, but differences between segments persist and helical regions are still identified. With the backbone correction of TIP4P-ws absolute helicities of TIP3P are reproduced or even exceeded. With TIP4P-D water folded states are energetically penalized and all peptides strongly favor unfolded conformations. In particular, it tends to underestimate the formation and persistence of secondary structure elements and seems to destabilize salt bridges of charged side chains, as seen for the Arg-Glu pairs of segments 401 and 411.

The difference in the radii of gyration in Figure 3.5 seems small but is of importance, as the systems were forced to cover all areas of the R reaction coordinate, which strongly correlates with R_G. Since the sampling is forced to all regions of R_G, TIP4P-D must consequently favor larger R_G. The same argument goes for the number of H-bonds. At low R peptides are forced into completely helical structures already featuring 7 H-Bonds, so all force fields do sample structures with a high count of H-bonds. Yet only TIP3P samples collapsed structures with an even higher number of hydrogen bonds to a relevant degree.

Best et al.[228] tested their new water model with the amber03w protein force field, but in the supplement provided an adaption for amber99sb*-ILDN used

in this work. The adapted amber99sb*-ILDN with TIP4P-ws overstabilized helices at higher temperatures in a Ac-$(AAQAA)_3$-NH2 peptide, but correctly sampled helix propensities at 300 K. This is in agreement with our simulations, where TIP4P-s, with increased solute–solvent interactions but no adjustment of the protein backbone dihedral potentials, undersampled helical states, but TIP4P-ws with the backbone modifications increased helix propensity. Piana et al.[222] in their validations mostly used the amber99sb-ILDN force field. Their validation results should be valid with amber99sb*-ILDN used here, which is only adds a modification to improve helix–coil transitions[250]. An explicit test of the helix–coil equilibrium of short sequences was, however, not part of the original force field validation.

3.5. Conclusion

In this chapter we investigated the property of intrinsically disordered regions to contain transient helical population with MD simulations and NMR secondary chemical shifts. Our model system Axin-1 showed intrinsically disordered behavior but increased transient helical content in the binding regions of two binding partners. This result is of significant importance for understanding the function of IDP regions in proteins. Even a small preference for adopting conformations close to the bound structure can significantly modulate the binding capacity of a protein segment that could be a general basis for fine tuning the binding properties of IDP containing proteins. While simulations with the traditional TIP3P water model reproduced this trend, water models explicitly parametrized for IDPs underestimated the helical content. With corrections to backbone parameters amber99SBws was able to reproduce and even slightly overestimate the higher helical propensities of the binding regions. TIP4P-D, however, strongly disfavors collapsed, folded peptide conformations. Hence, the choice of the force field water model remains to be of critical importance for studying the properties of intrinsically disordered proteins.

4. Mechanism of pKID/KIX Association

The phosphorylated kinase-inducible domain (pKID) associates with the kinase interacting domain (KIX) via a coupled folding and binding mechanism. The pKID domain is intrinsically disordered when unbound and upon phosphorylation at Ser133 binds to the KIX domain adopting a well-defined kinked two-helix structure. In order to identify putative hot spot residues of binding that could serve as initial stable anchor we performed in silico alanine scanning free energy simulations. The simulations indicate that charged residues including the phosphorylated central Ser133 of pKID make significant contributions to binding. However, these are of slightly smaller magnitude compared to several hydrophobic side chains not defining a single dominant binding hot spot. Both continuous Molecular Dynamics (MD) simulations and free energy analysis demonstrate that phosphorylation significantly stabilizes the central kinked motif around Ser133 of pKID and shifts the conformational equilibrium towards the bound conformation already in the absence of KIX. This result supports a view that pKID/KIX association follows in part a conformational selection process. During a 1.5 µs explicit solvent MD simulation folding of pKID on the surface of KIX was observed after an initial contact at the bound position of the phosphorylation site was enforced following a sequential process of α_A helix association and a stepwise association and folding of the second α_B helix compatible with available experimental results.

The contents of this chapter have been published[189] in similar form.

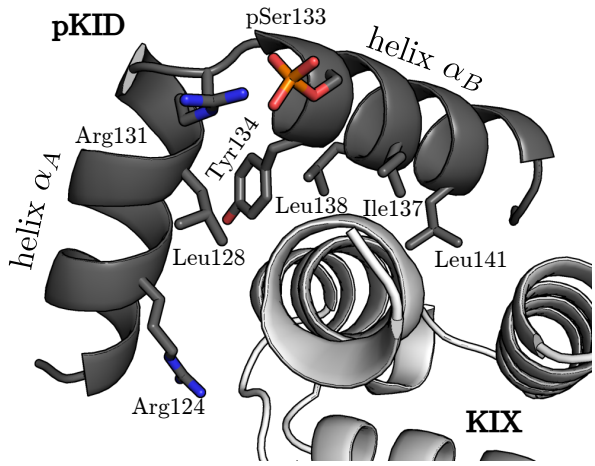


Figure 4.1.: pKID (gray cartoon) bound to KIX (white cartoon). Heavy atoms of side chains that face the binding interface and were subject to alchemical free energy simulations are shown in atom color-coded stick representation.

4.1. Introduction

Eukaryotic gene transcription is mediated by protein-protein interactions of DNA-bound factors and cofactors interacting with the basal transcription machines. Cyclic AMP response-element binding protein (CREB) and its co-activator CREB-binding protein (CBP) are essential for cell growth[262,263] and differentiation[264]. Also many human diseases, including leukemia, cancer and mental retardation[265], have been linked to CBP. CBP and CREB associate via their kinase-inducible domain interacting domain (KIX) and phosphorylated kinase-inducible domain (pKID), respectively. The isolated CBP-KIX domain forms a stable structure made of a small bundle of three interacting alpha-helices and two short 3_{10} helices joined by small loops enclosing a hydrophobic region. KID, however, is an intrinsically disordered protein (IDP) when unbound and only forms two stable helices, α_A and α_B, upon binding to KIX[266]. In the complex with KIX the two helices are oriented approximately perpendicular forming a 90 degree angle (Figure 4.1). Using nuclear magnetic resonance spectroscopy (NMR) it has been shown that the disordered KID still shows a residual helicity of 50% in helix α_A and 10% in helix α_B[266].

Phosphorylation of KID was shown to be of critical importance for CBP-

CREB binding[25,267] and increases the affinity by two orders of magnitude[268]. It is, however, still not completely clear how phosphorylation affects binding. Binding studies at different pH to affect the ionization state of the phosphate group showed a favorable electrostatic contribution[269]. Several direct contacts of the phosphate moiety with KIX residues Lys662 and Tyr658 observed in the experimental complex structure[25,268,270,271] have been identified and support an energetic stabilization of the complex. However, in addition a bias towards more helical structures in the free KID upon phosphorylation has been observed by an increase of helicity in helix α_B, observed via NMR-shifts[266]. Molecular dynamics (MD) simulations indicate a free energy landscape favoring the bound form upon phosphorylation[272]. Hence, phosphorylation may affect the equilibrium between folded bound conformation and unfolded (unbound) structures which indirectly influences the binding affinity of KID/pKID to KIX.

The process of coupled folding and binding can in principle take any path in between two extreme scenarios[273]. On the one side conformational selection is the process of exclusive binding of the correct, i.e. bound form from the unstructured ensemble. On the other hand in the induced fit mechanism any unstructured conformation can bind and transform into the bound structure after binding. NMR titration experiments[274] have identified an intermediate state of pKID on the surface when binding to KIX. Coarse-grained simulations[275] point in the same direction: Unstructured pKID binds to KIX and on the surface forms the bound structure. In principle MD simulations are ideally suited to investigate IDPs, since they can resolve structural dynamics of ensembles with accurate spatial and temporal resolution. Combined with free energy simulation methods it has been possible to calculate the absolute binding free energy of pKID and KID to KIX[272] and also to estimate a possible influence of the phosphorylation on the KID conformational ensemble.

In order to better understand the mechanism of coupled structure formation and binding in the present study we investigate the binding free energy contributions of individual interface residues. The characterization of hot spot binding residues may allow identification of stable initial binding placements that initiate the subsequent folding process. Interestingly, the simulations indicate that there is no single key residue in KID that dominates binding but several

hydrophobic residues and the phosphorylated Ser133 make similar binding free energy contributions. In addition, we use both extensive unrestrained MD simulations and free energy simulations to investigate the equilibrium between unbound (unfolded) and bound KID conformations and how it is influenced by phosphorylation of Ser133 in pKID. Although convergence of free energy estimates appears to be difficult to achieve the calculations support a view that phosphorylation indeed shifts the conformational equilibrium towards the bound form even in the absence of the KIX binding partner. The latter result also supports a view that pKID-KIX binding is not entirely based on induced-fit but at least in part also involves a phosphorylation controlled conformational selection mechanism. Finally, a folding simulation of disordered pKID initially bound at the pSer133 binding site results in rapid formation and near native binding of the α_A helix followed by stepwise formation and binding of the α_B helix. This agrees with the coupled binding and folding model suggested by NMR spectroscopy[274] and allows obtaining atomic level insight into the binding mechanism.

4.2. Methods

All the simulations were performed using GROMACS 4.6.5[251] with periodic boundary conditions and the Particle-Mesh-Ewald[136] method with a Fourier-spacing of 0.16 nm and a grid interpolation up to order 4 to account for long-range electrostatic interactions. Close Coulomb real space interactions were cut off at 1.0 nm and Van-der-Waals interactions were cut off after 1.4 nm. Long range dispersion correction was applied to account for errors stemming from truncated Lennard-Jones interactions. To constrain bond lengths of H-atoms LINCS[253] was used with a coupling matrix extension order of 4 (12 in equilibration runs). All simulations except the association simulation were started from equilibrated conformations of the NMR structure (model 1 in pdb 1KDX).

Initial energy minimization was performed with the steepest descent algorithm for 50000 steps. For each replica or continuous simulation a NVT equilibration for 50 ps, controlling the temperature at 300 K with velocity

rescaling[276], and a NPT equilibration of 100 ps using the Berendsen[277] barostat were run with a step size of 1 fs. For the production runs the time step was increased to 2 fs and the Parrinello-Rahman barostat[125] was used. All production simulations were performed at a temperature of 300 K and a pressure of 1.01 bar.

In Silico Alanine Scanning

In silico free energy perturbation (FEP) simulations, evaluated with the Bennett Acceptance Ratio[278] method (FEP/BAR), were performed with the amber99sb-ILDN[279] force field with TIP3P[280] water and phosphoserine parameters from Homeyer et al.[281]. Along an alchemical transformation parameter λ 21 replicas with evenly spaced λ values were used. To improve convergence "Soft-core" potentials[282,283] were applied both for Lennard-Jones- and coulombic potentials with $\alpha = 0.3$, linear lambda scaling and a radius power of 6. For changes of charge we separated mutation of charge- and vdW-interactions[284] and 3 simulations with 21 λ windows were conducted: For both complex and ligand first the charges of the wild-type side chain were removed, then the uncharged wild-type side chain was transformed into the uncharged mutant and finally the mutant's side chain charges were turned on again. Separate FEP/BAR simulations for the complex and the isolated pKID protein were performed. To minimize effects of charge alteration due to the finite system size equal box sizes were used for the complex and the ligand. Free energy differences were calculated using the Bennett Acceptance Ratio method as implemented in GROMACS 5.0.5. $\Delta\Delta G$-values were also calculated using thermodynamic integration, which yielded similar results, though with larger errors.

Available experimental data from ITC experiments[267,268] had been gathered at 50 mM salt concentration while for the simulations we only added salt to neutralize the system, which was approximated as a 10 mM concentration. To correct for this discrepancy the electrostatic effect of a higher salt concentration was estimated. Electrostatic energies were calculated by solving the Poisson-Boltzmann equation with the adaptive Poisson-Boltzmann-solver APBS[285]. The change of the binding free energy difference, $\Delta\Delta G_{50\,\mathrm{mM}}$ - $\Delta\Delta G_{10\,\mathrm{mM}}$ can be

estimated as

$$\Delta\Delta G_{50\,\mathrm{mM}} - \Delta\Delta G_{10\,\mathrm{mM}} = (\Delta G^{\mathrm{mut}}_{50\,\mathrm{mM}} - \Delta G^{\mathrm{mut}}_{10\,\mathrm{mM}}) - (\Delta G^{\mathrm{wt}}_{50\,\mathrm{mM}} - \Delta G^{\mathrm{wt}}_{10\,\mathrm{mM}}) \quad (4.1)$$

where ΔG^{x}_c is the electrostatic contribution to binding at salt c for the wild type (wt) or the mutation (mut), calculated as the difference of APBS electrostatic energies: $\Delta G^{\mathrm{x}}_c = E^{\mathrm{complex}}_c - (E^{\mathrm{(p)KID}}_c + E^{\mathrm{KIX}}_c)$. The final correction applied (see Table 4.1) was an average over complex, KIX and (p)KID structures taken from 10 snapshots of the 200 ns simulation of the complex. Identical structures were used for both wild type and mutated systems, apart from the mutation.

Conformational Analysis

Continuous MD simulations of free pKID and KID were performed for 1040 ns and employed the amber99sb*-ILDN[279,286,287] force field with phosphoserine parameters[281] as before but the TIP4P-D[222] water force field optimized for intrinsically disordered proteins. Test simulations indicated that a uniform cutoff (1.1 nm) for real space electrostatic and Lennard Jones interactions resulted in essentially the same dynamics and RMSD distribution as using two different cutoffs (see above). For greater computational efficiency the uniform cutoff scheme was used for the simulations of the free (p)KID.

To sample the full conformational space of the unbound and unfolded pKID as a reaction coordinate R we used the root mean square deviation of a set of N distances d_i relative to the distances in the bound form, $d_{i,0}$:

$$R(d_{1..N}) = \sqrt{\frac{1}{N}\sum_i^N (d_i - d_{i,0})^2} \quad (4.2)$$

Harmonic potentials forced the systems to specific regions along R. Replica Exchange allowed overcoming artificial energy barriers within umbrella windows.

Replica Exchange dRMSD umbrella sampling simulations were run with 32 replicas for 20 ns for the full pKID and 200 ns for the core motif, attempting exchanges every 1 ps. All replicas were started at the equilibrated bound

structure. The unfolding coordinate R was sampled between 0.0 - 0.5 nm for the full protein and between 0.0 - 0.8 nm for the core motif, in evenly spaced λ-intervals enforced by harmonic potentials with a force constant of 400 kJ/mol/nm^2. PMFs along the R were obtained using the WHAM[288,289] algorithm.

Association Simulations

To follow the structure formation process of pKID at the binding surface of KIX 10 simulations starting from an unfolded pKID structure (snapshot taken from the above simulations of the isolated pKID) employed the same force fields as the alanine scanning simulations and were run for 500 ns. During the simulations the C_α atoms of KIX were weakly restraint to positions in the bound starting structure with a force constant of 250 kJ/mol/nm^2. To pull the disordered pKID structure near the binding position on the KIX surface a positional restraint of the C_α atom of pSer133 with respect to the position in the complex and using a force constant of 50 kJ/mol/nm^2 was included. The restraint enforced an initial encounter at the binding site of pSer133 and avoided dissociation during the subsequent folding process. During the 500 ns simulation time only one simulation showed a significant decrease of the deviation from the bound structure and signs of folding on the surface and was continued up to 1.5 µs.

4.3. Results

During a 200 ns simulation of the pKID-KIX complex starting from the experimental structure the protein partners stayed in a stably bound conformation with a root mean square deviation (RMSD) from the starting structure of < 4 Å, as shown in Figure 4.2. This proves, that in our simulations and with our force field the bound state is a stable free energy minimum.

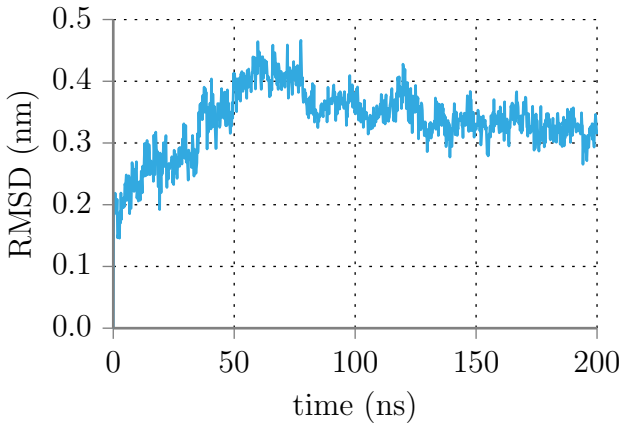


Figure 4.2.: Backbone RMSD of pKID-KIX complex simulation relative to the NMR structure (pdb1KDX, model 1) during MD simulation. Deviations mostly arise from a slight reorganization of helix α_A.

4.3.1. Contributions of Side Chains to Binding Free Energy

The binding interface of the pKID/KIX complex consists of several non-polar side chains supplemented by charged and polar residues at the rim region of the interface (see Figure 4.1). As suggested by Sugase and coworkers[274] based on results of NMR spectroscopy a possible mechanism of coupled folding and binding of pKID may involve an initial encounter contact of a pKID residue with the interface region followed by structure formation on the KIX surface. The initial anchoring must be stable enough for a sufficiently long time to result in productive pKID folding and formation of additional contacts without prior dissociation. In order to identify such putative hot spot binding residues we performed in silico free energy simulations to calculate the binding free energy contribution of interface side chains by alchemical transformation to Alanine (Ala). The contribution was obtained by performing the corresponding transitions in the complex and in the isolated binding partner in 21 steps (3×21 steps for charged side chains) of a coupling parameter lambda (see Methods for details). In order to check the accuracy of the approach in silico "mutations" for which experimental data is available were included.

The calculated differences in the free energy of binding for several mutations, shown in Figure 4.3 and Table 4.1, show very good agreement with experimen-

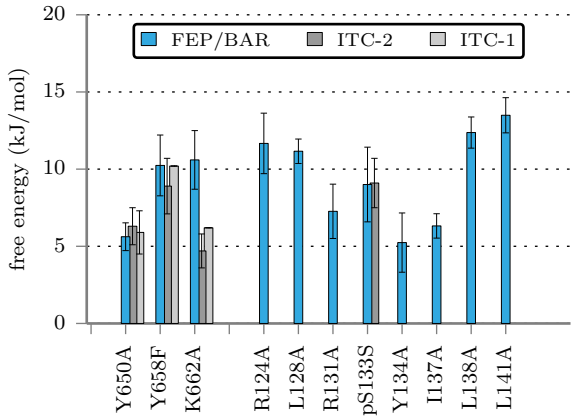


Figure 4.3.: Calculated free energy contributions to binding of side chains at the pKID/KIX interface using the FEP/BAR method. Available experimental data are shown as dark and light gray bars from two separately published measurements: ITC-1[267] and ITC-2[268].

.

tal values for uncharged side chains and reasonable agreement for charged side chains. The free energy contribution of all three pKID Leu interface residues is similar (11 - 14 kJ/mol) and slightly larger than the calculated contribution of Arg124 and of the phosphoryl group attached to Ser133. The most important contributions arises from hydrophobic contacts in helix α_B, but also contacts in helix α_A and the phosphorylation at Ser133 contribute significantly.

A closer look at the trajectories indicates the detailed interactions in the complex, illustrated in Figure 4.4): Of the KIX side chains Tyr650 occasionally establishes a hydrogen bond to the pKID backbone oxygen at Ala145 and seals off the hydrophobic core around Leu141, whereas water entered the interface upon Tyr650Ala mutation. The hydroxyl group of Tyr658 and the amine of Lys662 are constantly contacting the phosphate moiety, forming crucial links which are missing when mutated to Phe or Ala, respectively. For pKID the most important contributions to binding arise from the hydrophobic contact of residues Ile137, Leu138 and Leu141 with the hydrophobic groove of KIX which tightly bound in the interface of the complex. In the other helix α_A Arg124 either forms salt bridges with Glu655 or Asp120, while in

	$\Delta\Delta G$	ITC 1[25]	ITC 2[268]	decharge	vdw	recharge	PB
Y650A	5.62	5.90	6.30				
Y658F	10.24	10.20	8.90				
K662A	9.98	6.20	4.70	9.74	0.10	0.14	-0.62
R124A	10.83			11.89	-1.05	-0.01	-0.84
L128A	11.16						
R131A	6.24			6.03	0.36	-0.15	-1.03
pS133S	10.51		9.10	10.00	2.91	-2.40	1.51
Y134A	5.24						
I137A	6.32						
L138A	12.37						
L141A	13.49						

Table 4.1.: Data of free energy perturbation simulations and difference of Poisson-Boltzmann-calculations for salt concentrations 50 - 10 mM. All values in kJ/mol.

simulations of the free pKID Arg124 also connects to the C-terminal end of pKID. Leu128 in the complex contacts the phenyl rings of Tyr650 and Tyr134. In the kink region Arg131 is highly solvent accessible in the complex but also tightly bound to pSer133 in both the complex and during simulations in the absence of the binding partner. Tyr134 establishes a stable H-Bond with His651 in the complex. Finally, pSer133 tightly binds to Tyr658 and Lys662 in the complex simulation and forms two stable salt bridges with Arg131 in both simulations. When bound pKID remained in a stable folded structure for all in silico mutations. Free energy simulations in the absence of the binding partner show signs of unfolding but the simulation time was insufficient to sample major conformational changes. The alchemical transformations do not indicate a single dominant hot spot binding residue in pKID that might be responsible for an initial stable anchoring of the unstructured peptide.

4.3.2. Unrestrained MD Simulations of KID and pKID

The influence of individual side chains on the helix-coil transition of peptides is in most cases below 1 - 2 kJ/mol[290]. The contribution of individual side chains to the equilibrium between conformations close to the bound pKID structure and unfolded conformations is therefore likely to be much smaller than the

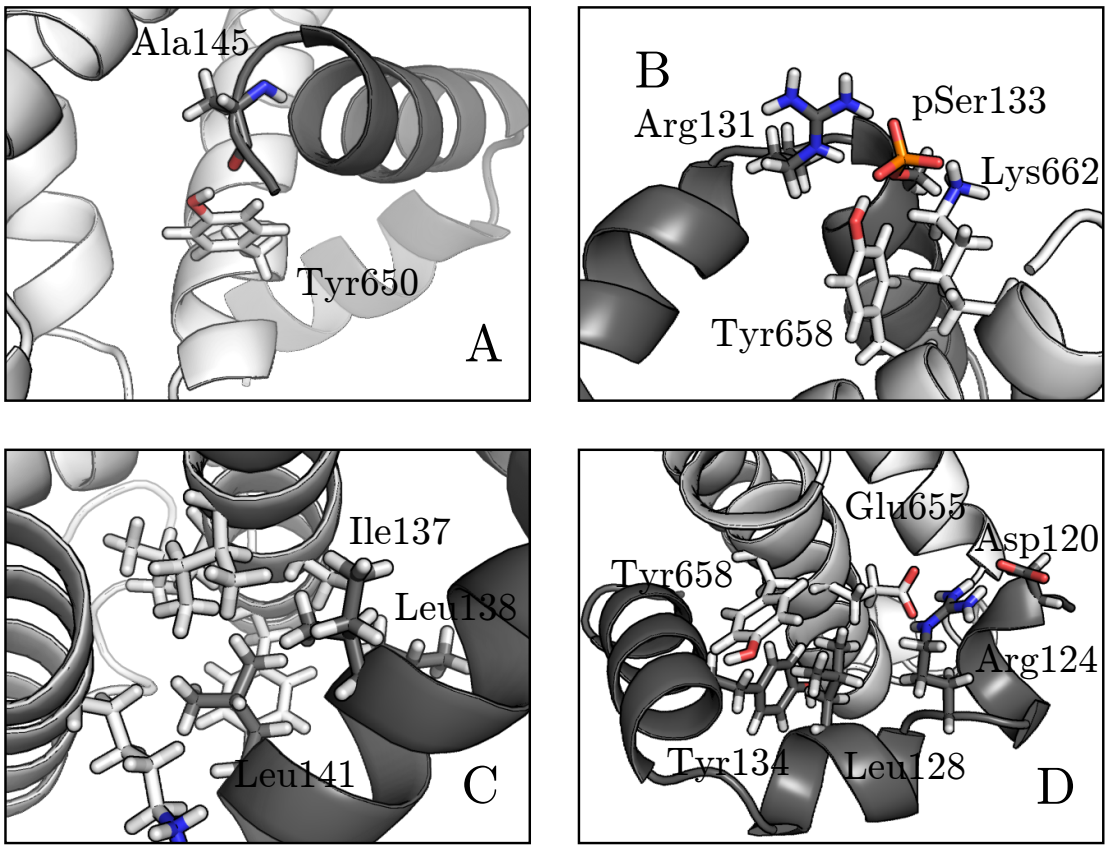


Figure 4.4.: Snapshots from wild type replicas of the FEP-BAR simulations. Image A shows the hydrogen bond between Tyr650 and Ala145, B shows the interaction network around the phosphorylation site, C shows side chains of pKID's helix α_B in contact with the hydrophobic groove of KIX and D shows interactions of pKID's helix α_A.

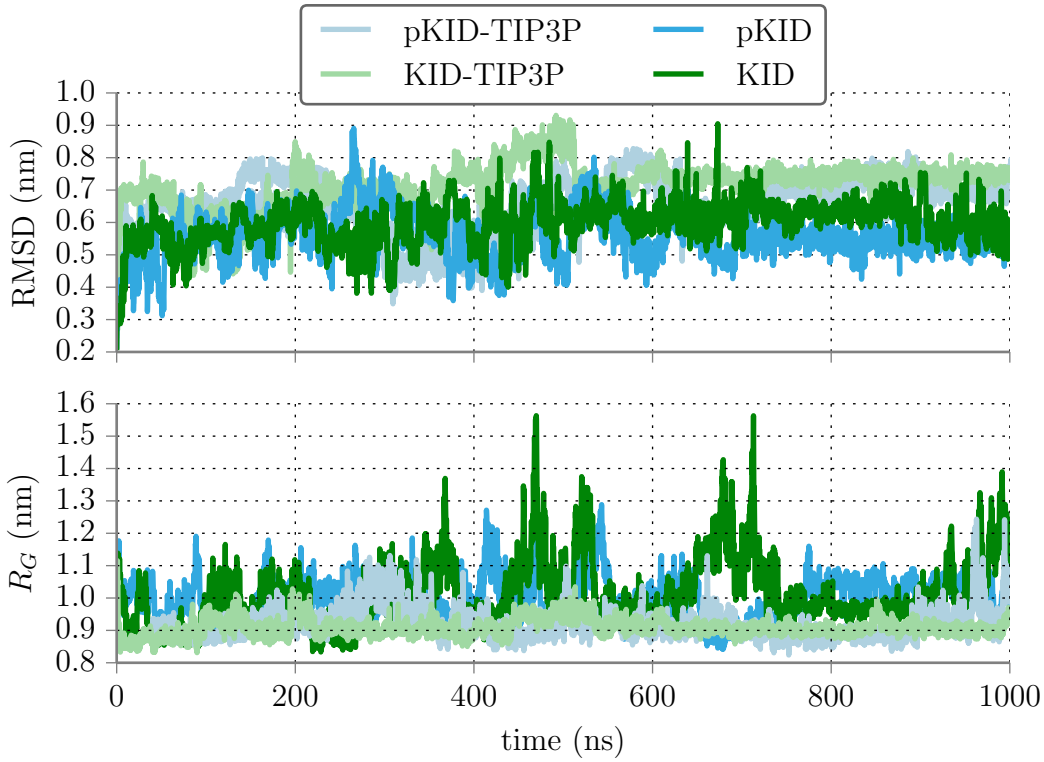


Figure 4.5.: Backbone RMSD and radius of gyration of continuous MD simulations with TIP3P and TIP4P-D water models. With TIP3P structures can be stuck for hundreds of ns, e.g. for KID from 600 - 1000 ns. R_G shows with TIP3P only collapsed states are sampled.

above calculated free energy contributions to binding. However, for the removal (or creation) of a phosphoryl group at Ser133 this may not be negligible, since it involves removal of two unit charges involved in long-range electrostatic interactions. In the particular case of pKID it furthermore involves the removal of a specific structural kink motif formed with Arg131 that may stabilize the bound conformation. To assess the behavior of unbound phosphorylated and non-phosphorylated KID we performed 1 µs continuous MD (cMD) simulations with TIP4P-D[222] water parameters specifically developed to cover the dynamics of intrinsically disordered proteins. For comparison simulations were also performed with TIP3P water. Results are shown in Figure 4.5.

During 1.1 µs simulations both pKID and KID largely unfold. For KID helix α_B completely unfolds at times and α_A unfolds from the N-terminus and

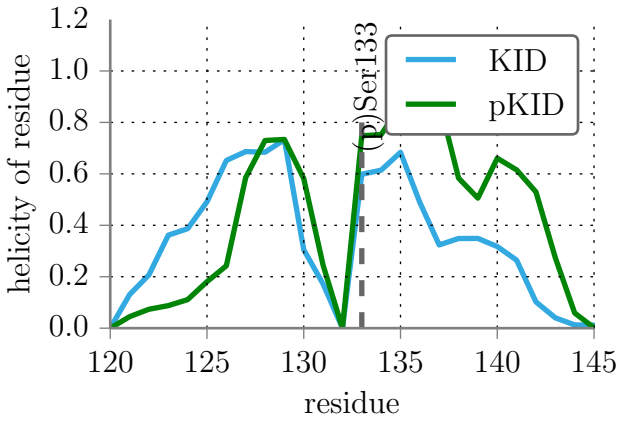


Figure 4.6.: Mean helicity calculated with the DSSP algorithm[260] for 1 µs continuous simulations of pKID and KID, respectively.

occasionally reforms a helical turn. pKID, in contrast, unfolds from the ends but maintains helicity close to the phosphorylation site for both helices. Helicities according to the DSSP algorithm[260] in the two helical regions α_A (residues 120-129) and α_B (residues 134-144) are 46% and 27% for KID and 34% and 54% for pKID, respectively. KID reproduces the trend from NMR-measurements[266], where 50-60% for helix α_A and 10% for α_B were estimated. For pKID our strong observed increase of helical content in helix α_B exceeds the NMR-estimates of 15%, while no decrease in helix α_A was observed in experiment. In particular, phosphorylation further increases the higher helical content in the core around the phosphorylation site itself (Figure 4.6). A coupling of helicity and deviation from the experimental structure of the sampled pKID and KID structures was observed, shown in Figure 4.7.

On the atomic scale the pSer133 group strongly interacts with the charges of Arg131 and, to a lesser degree, of Lys136. This, together with the formation of a stable hydrophobic core of residues Ile121, Leu128, Tyr134(phenyl ring), Ile137 and Leu138, leads to a stabilization of the opening angle, as shown in Figure 4.8. For KID, in contrast, the angle frequently opens up and no stable central kinked motif is observed.

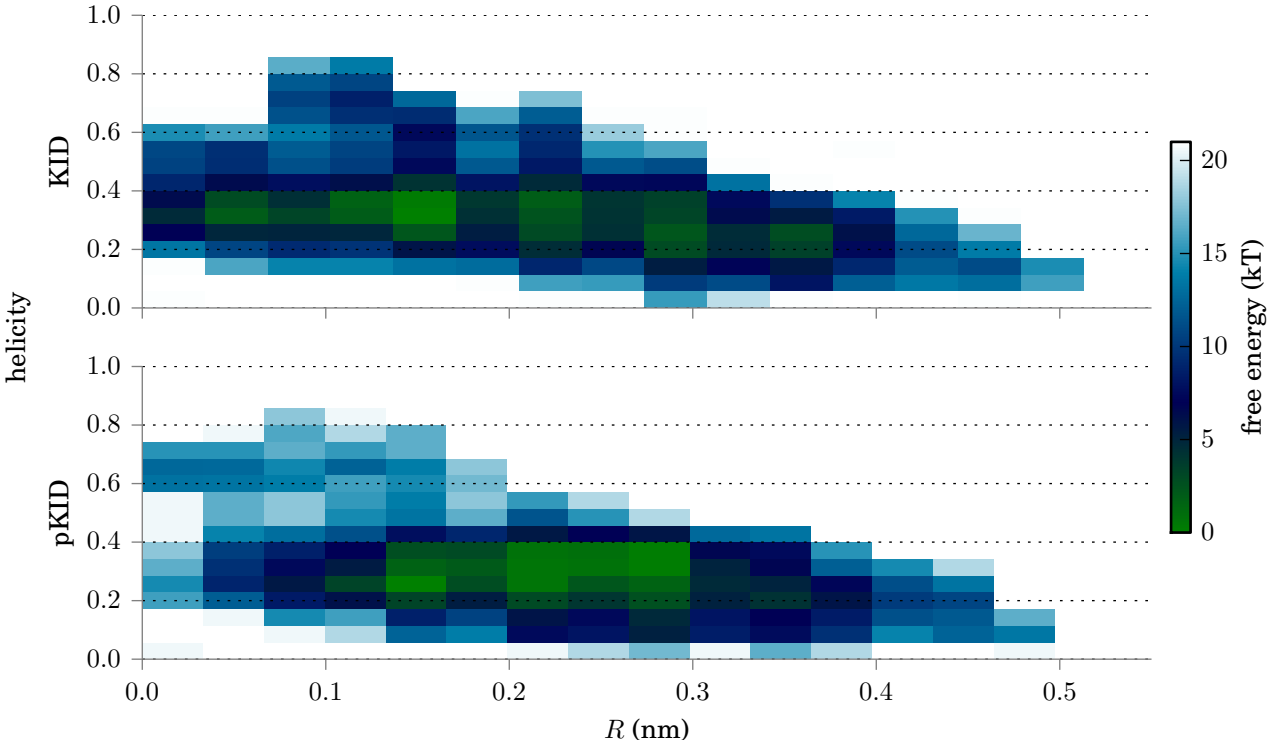


Figure 4.7.: Free energy surface in the space of helicity and dRMSD reaction coordinate R (as defined for the core motif) for continuous MD simulations of and KID and pKID. The main free energy basin of pKID is narrower and at slightly higher helical content. KID structures are distributed over a wider area of R but also reach higher helical content more frequently.

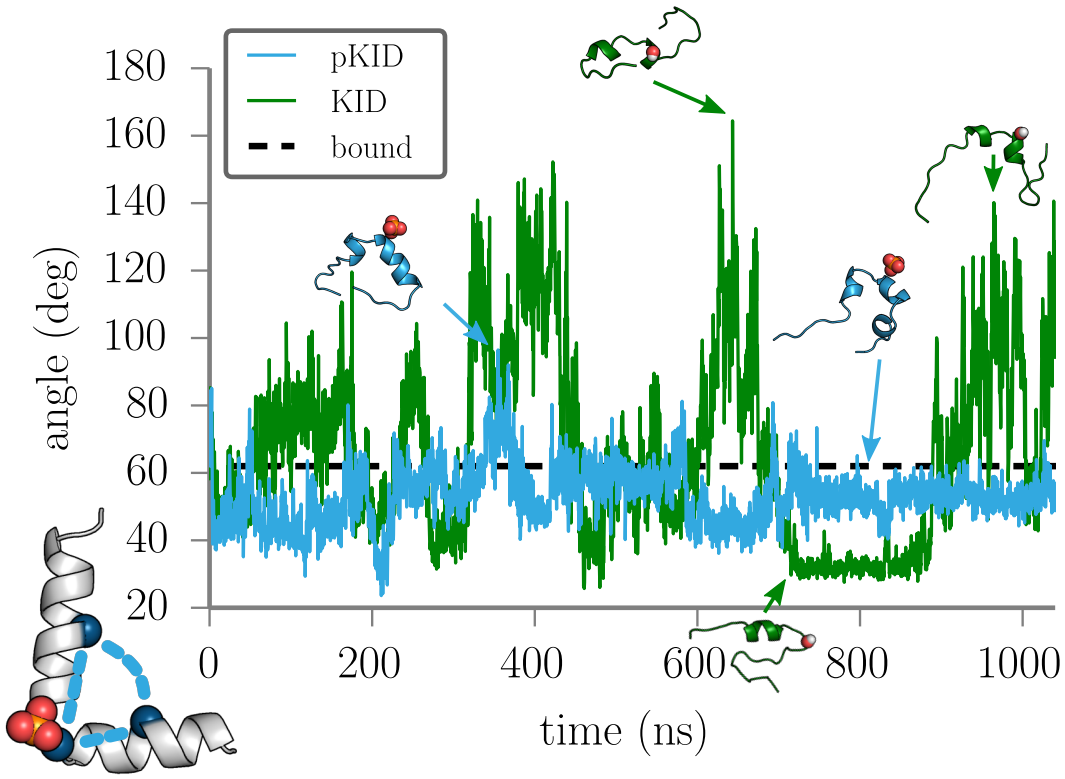


Figure 4.8.: The angle of the C_α atoms of residues 125, 131 and 138 of pKID and KID, illustrated for the white cartoon structure, served as an indicator of the global shape and has been recorded for the 1 µs simulations. Values near the bound form (dashed line) indicate a geometry of the central kink motif similar to the bound structure whereas large deviations indicate a disruption of the motif. Typical sampled snapshots are shown as cartoons along the plot.

4.3.3. Free Energy of Unfolding Along the dRMSD

To control the trends obtained from the cMD simulations Replica Exchange Umbrella Sampling (REUS) simulations along the mean deviation of a number of distances, the dRMSD R, were performed. A dRMSD of $R = 0\,\text{nm}$ indicates a structure in exact agreement with the native bound conformation whereas a dRMSD of $R = 0.3 - 1.0\,\text{nm}$ corresponds to unfolded structures. The simulation of the full pKID or KID with 32 replicas, each simulated for 20 ns data gathering time, indicated a broad minimum around $R = 0.2\,\text{nm}$ for both pKID and KID (Figure 4.9). However, the increase in R along the reaction coordinate was mainly the result of unfolding the last helical turns of both the N- and C-termini of the peptide. For this part of the transition a similar shape near the free energy minimum was found for both pKID and KID. Hence, not surprisingly, the unfolding of the terminal helical turns is not affected by the phosphorylation of the central Ser133. For KID, however, two secondary minima at the unfolded end of the R scale are observed coupled to an onset of unfolding of the central core motif. Thus, differences in the stability of pKID vs. KID arise near the central kink motif, in qualitative agreement with the cMD simulation results of the previous paragraph.

The conformational flexibility of the central motif is poorly sampled in dRMSD free energy simulations of the complete pKID or KID protein. A simulation time of 20 ns for each of 32 windows, already far beyond the simulation time of previous studies to characterize the stability of KID or pKID[272], was found to be still insufficient for a converged PMF (Figure 4.9). Since the most significant changes upon phosphorylation seem to affect mainly the central structural motif we conducted further dRMSD-REUS-simulations biasing solely the core motif structure. Only pairs of C_α-atoms in residue range 125-138, less than 0.7 nm apart and with at least 3 residues between them in the peptide sequence were included (see Figure 4.10). Even for this much smaller peptide segment convergence of the folding free energies along the dRMSD reaction coordinate was difficult to achieve, requiring 200 ns per US window and shown in Figure 4.11. Note, that it is expected that for the full pKID or KID even longer sampling per window is required. The global minimum of the central KID motif clearly shows that KID favors an open, unfolded motif

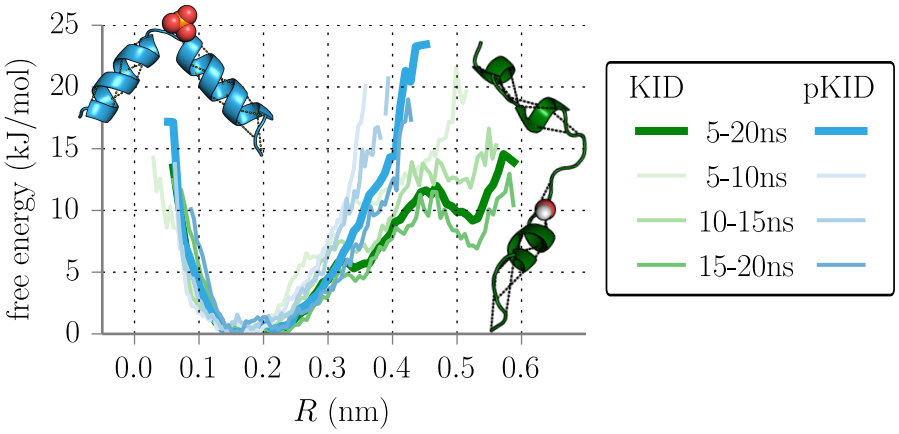


Figure 4.9.: Calculated PMF of the full KID protein unfolding/folding along the R coordinate where low dRMSD indicates close resemblance to the folded structure and high values correspond to unfolded structures. Thick lines indicate the total PMF after 5 ns equilibration, thinner lines show PMFs of simulation time intervals of the simulation. Full (p)KID, as shown in the structure inset, was simulated with all pairs contributing to R indicated by yellow dashed lines.

conformation whereas the phosphorylation of Ser133 shifts this minimum to a folded conformation relatively close to the native bound pKID structure (Figure 4.10). The free energy difference of being in the folded state, here defined as $R < 0.2$ nm, is 11.3 kJ/mol lower for the phosphorylated peptide in comparison with the non-phosphorylated KID.

4.3.4. Simulation of the pKID Folding Process

To directly follow the encounter and subsequent folding of pKID near the binding site on KIX we initiated 10 simulations started with an unstructured and unbound pKID extracted from the simulation of the free pKID and dragged to the binding site by a single positional restraint on the C_α atom of pSer133. We note that while in principle initial binding can occur at several positions (see previous paragraphs) it is reasonable to assume that long range electrostatic interactions of pSer133 may guide pKID towards the initial approximate binding position. Out of 10 simulations 9 did not get any closer towards the bound structure within the first 500 ns and were terminated (data not shown).

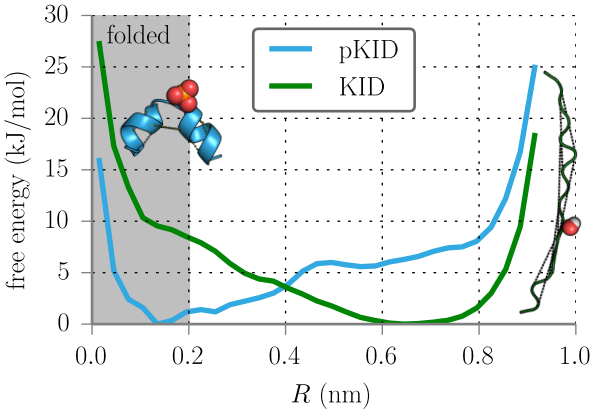


Figure 4.10.: Calculated potential-of-mean-force (PMF) along R for the kinked core motif of (p)KID, as shown in the inset structure. Pairs contributing to R are shown as yellow dashed lines in the depicted structures.

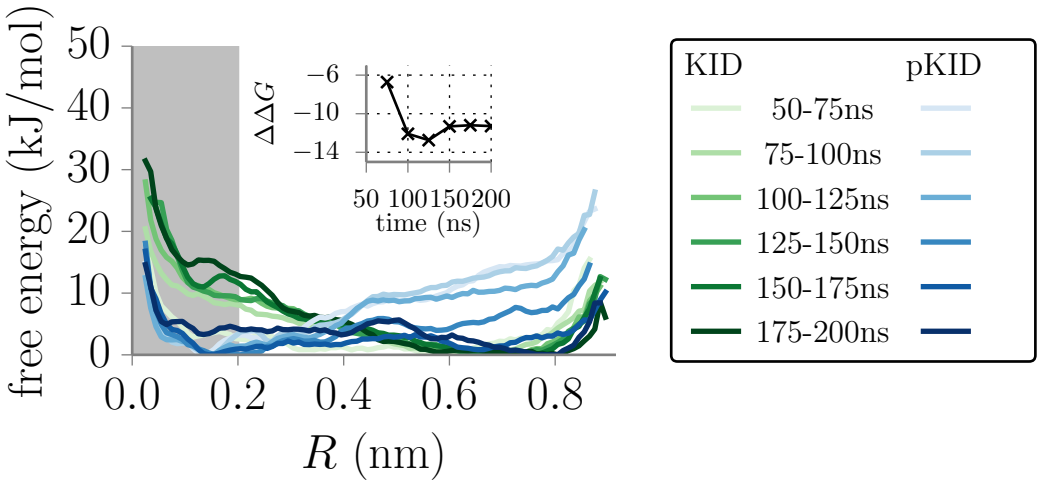


Figure 4.11.: Calculated PMF obtained from different time intervals of the US simulations along the dRMSD coordinate of the kinked core motif. Inset shows the difference of free energy gain upon folding: ΔG_{fold}(KID) - ΔG_{fold}(pKID). For the different time intervals the calculated free energy difference between pKID and KID remains approximately constant.

In one trajectory, however, the helix α_A folded and moved to a near-native position on the KIX surface after 100 ns (Figure 4.12). Such an intermediate initial binding of the α_A helix mediated by residues 124-128 was also observed in the NMR study of Sugase et al.[274] and the trajectory was continued up to 1.5 µs. Subsequently, at ∼230 ns, the first turn of helix α_B close to pSer133 formed. This was accompanied by a significant change of the global kink angle between helices towards the bound form, displayed in Figure 4.13. The remainder of helix α_B then was locked on the surface of KIX for 1 µs, partially detaching twice. The second detachment allowed a twist of the C-terminal end and the intrusion of Leu141 into the binding pocket, effectively locking the bound position. Interestingly, also the NMR study of Sugase et al. found that the α_B helix is incompletely folded in the encounter and intermediate stages of binding and proposed two separate binding regions in the α_B region that bound in a stepwise manner during the coupled binding and folding process.

4.4. Discussion

The in-silico alanine scanning of pKID indicates that, although the pSer133 and also some charged residues contributed significantly to binding to KIX, the largest contributions arise from hydrophobic contacts. However, a preferable initial hot spot contact residue could not be identified. Phosphorylation of Ser133 was found to have a two-fold effect: The phosphate moiety directly interacts favorably with KIX through hydrogen bonds with Lys662 and Tyr658 of KIX, but also alters the conformational free energy landscape in favor of the folded structure. Although previous experimental and simulation studies (of much shorter length) came to similar conclusions our simulations clearly indicate that the central kinked structural motif involving specific contacts between pSer133 and its neighboring residues is responsible for this effect. The phosphorous group contributes directly to binding through hydrogen bonds and reduces the overall charge of the positively charged central segment. As suggested earlier[272,291] this stabilizes a compact structure for this segment. The present continuous MD simulations also indicate that isolated KID undergoes larger global fluctuations than pKID.

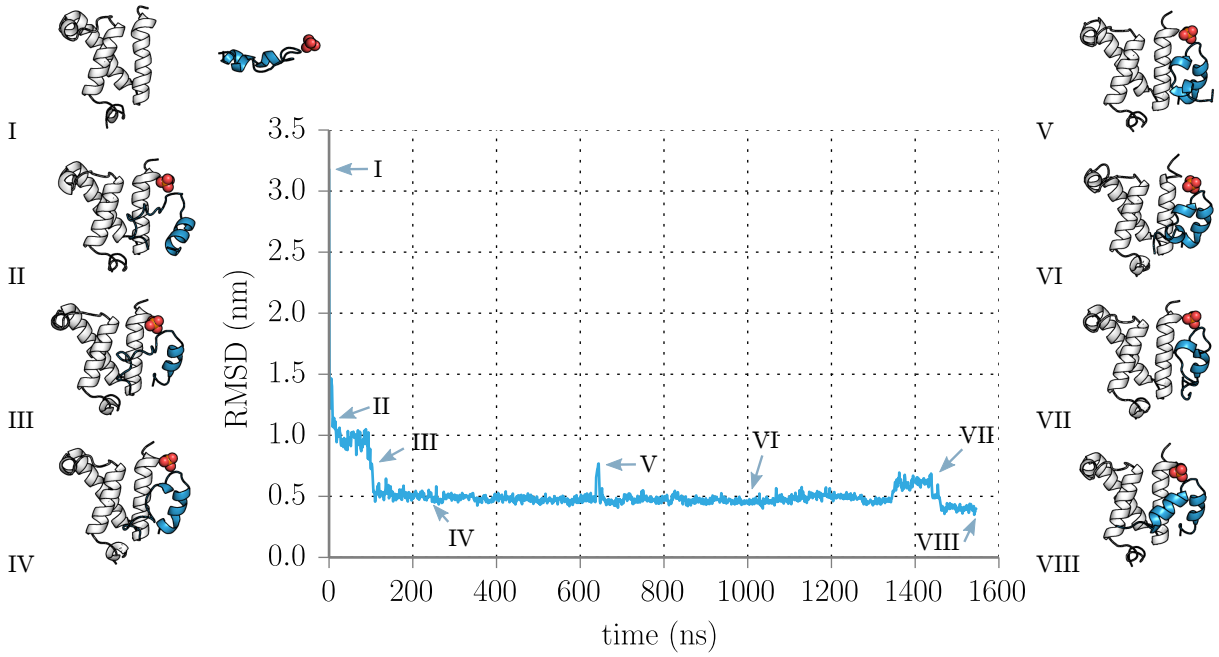


Figure 4.12.: RMSD of association of pKID with respect to the NMR structure (pdb 1KDX model 1). RMSD is calculated for backbone atoms of pKID after a fit of KIX backbone atoms. When calculated with respect to the final MD equilibrated complex structure the RMSD exhibits similar behavior but is lowered by approximately 1 Å. Images depict crucial events of coupled folding and binding. I: Starting structure with unfolded and detached pKID. II: pKID pulled to binding site via harmonic restraint on $C_{\alpha,133}$ III: Helix α_A flips to correct position. IV: First turn of helix α_B forms on surface. V: α_B region detaches. VI: Semi-stable structure with α_B region incorrectly folded. VII: Second detachment of α_B region. VIII: Correct folding with insertion of Leu141.

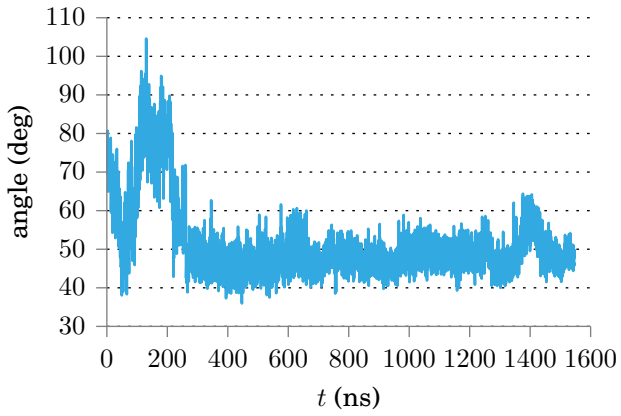


Figure 4.13.: Angle between C_α atoms of residues 125, 131 and 137 sampled during the successful association simulation of pKID to KIX. The correct folding of helix α_A (at 200 ns) and the first helical turn of helix α_B (after 240 ns) is coupled to a drop of the kink angle.

The calculated relative free energy difference of binding to KIX for pKID compared to KID consists of a direct contribution (obtained from the alchemical transformation) and an indirect contribution due to a relative stabilization of the bound form which sums up to $(-10.51 - 11.3)\text{kJ/mol} = -21.81\,\text{kJ/mol}$. It agrees with other simulation results ($-19.2\,\text{kJ/mol}$[272]) but exceeds experimental results of $-9.1\,\text{kJ/mol}$[268] or $-7.32\,\text{kJ/mol}$[292]. Both contributions, obtained from alchemical transformation of pSer133 to Ser133 ($-10.51\,\text{kJ/mol}$) and stabilization of the folded form due to phosphorylation ($-11.3\,\text{kJ/mol}$), also roughly agree with the simulation results of Dadarlat & Skeel's[272], with $-7.07\,\text{kJ/mol}$ for the direct interactions with KIX and $-15.3\,\text{kJ/mol}$ for the conformational contribution. However, it should be emphasized that simulation times of 200 ns for each of the 32 umbrella windows for unfolding were required to achieve reasonable convergence of the calculated free energies of unfolding which is considerably longer than previous studies using 3 ns each for 18 umbrella windows[272].

The pKID folding studies on the surface of KIX upon initial anchoring of the disordered pKID start structure near the pSer133 binding site resulted in successful structure formation within 1.5 µs in one case. Although other simulations were unsuccessful and it is not possible to extract statistics on the

mechanism of folding the observed folding process agrees remarkably well with experimental findings of Sugase & colleagues[274]. Based on NMR results these authors proposed a three step mechanism based on the formation of an encounter complex followed by an intermediate complex and formation of the final bound state. In the intermediate state the α_A helix of pKID was found to be already formed and bound in a near-native geometry[274]. Such a state was also observed in our simulations as first intermediate structure after the initial encounter mediated by the pSer133 binding. The subsequent process can be roughly considered as a two-step process taking longer than formation of the first intermediate state and resulting in a partially formed α_B helix bound to KIX followed by a final folding process to result in the completely folded pKID and bound complex.

4.5. Conclusion

The results of this study support a coupled folding and binding mechanism of pKID to KIX that is a mixture of conformational selection and induced fit. The free energy simulations suggest that at least the central structural motif involving pSer133 and also the α_A helix are at least transiently already formed in the absence of the binding partner and allow initial binding in part based on conformational selection (or very rapid formation of the α_A helix after the encounter). The subsequent folding of the α_B helix can involve several steps and is driven by the significant free energy contributions of hydrophobic side chains that become buried at the interface.

5. 4E-BP2 Protein Fold Stabilization Induced by Phosphorylation

Protein phosphorylation can affect the interaction with partner proteins but can also induce conformational transitions. In case of the eukaryotic translation initiation factor 4E-binding protein 2 (4E-BP2) threonine (Thr) phosphorylation at two turn motifs results in transition from a disordered to a folded structure. In order to elucidate the stabilizing mechanism we employed comparative Molecular Dynamics (MD) free energy simulations on the turn motifs indicating that Thr-phosphorylation favors a folded whereas dephosphorylation or substitution by Glu residues destabilizes the turn structure. In multiple unrestrained MD simulations at elevated temperature of the 4E-BP2 domain only the double phosphorylated variant remained close to the folded structure in agreement with experiment. Three surface Arg residues were identified as additional key elements for the tertiary structure stabilization of the whole phosphorylated domain. In addition to the local turn structure double phosphorylation also leads to an overall electrostatic stabilization of the folded form compared to wild type and other investigated variants of 4E-BP2. The principles of phosphorylation mediated fold stabilization identified in the present study may also be helpful for identifying other structural motifs that can be affected by phosphorylation or provide a route to design such motifs.

The contents of this chapter have been published[293] in similar form.

5.1. Introduction

The impressive level of versatility of the cellular machinery is, to a significant part, facilitated by post-translational modifications (PTMs). The highly dynamic manner of attachment and removal of such PTMs allows for fast and precise activation and deactivation of protein functions. In eukaryotes, regulation via PTMs is used in practically every aspect of cellular activity, from cell growth and differentiation to programmed cell death[294]. Of the many different possibilities[89] to modify proteins after transcription, phosphorylation, i.e. the attachment of a phosphate group to a hydroxyl group of a protein side chain, is the most abundant[295]. Phosphorylation at Tyr, Ser or Thr side chains exerts its regulatory function often by influencing the binding properties of a peptide or protein to a partner if located at or close to the binding interface. However, phosphorylation has also been shown to influence the structural properties of proteins[294,296–298], capable of inducing both disorder[104] and order[105,106], e.g. by stabilizing α-helices[107,108], in proteins.

In a recent study, multiple site phosphorylation of the eukaryotic translation initiation factor 4E-binding protein 2 (4E-BP2) was found to not only affect the conformation of secondary structure elements but to induce folding of an entire 4-stranded β-sheet domain[299]. 4E-BP2 plays a critical role during cap-dependent translation initiation by binding to eukaryotic translation initiation factor 4E (eIF4E), which prevents translation initiation and is vital for synaptic plasticity and learning[300]. 4E-BP2 is the major eIF4E binding partner in the brain[301] and binds to eIF4E via a YXXXXLΦ-motif, which forms an α-helix when bound. Dissociation rates of 4E-BP2 from eIF4E are increased by two orders of magnitude when two sites, Thr37 and Thr46, are phosphorylated and can even increase by four orders of magnitude for a 5-fold phosphorylation of a larger segment of 4E-BP2. Thus, phosphorylation effectively prevents binding to eIF4E. Wild type 4E-BP2 is intrinsically disordered[302,303], but phosphorylation at Thr37 and Thr46 leads to the formation of a stable fold of residues 18-62 in a structure with 4 β-sheets and a short 3_{10}-helix, covering the binding motif and effectively locking the tyrosine of the binding motif in a hydrophobic core (see Figure 5.1) making it inaccessible for partner binding. The two phosphorylated Threonines Thr37 and Thr46 are embedded in identical pTPGGT

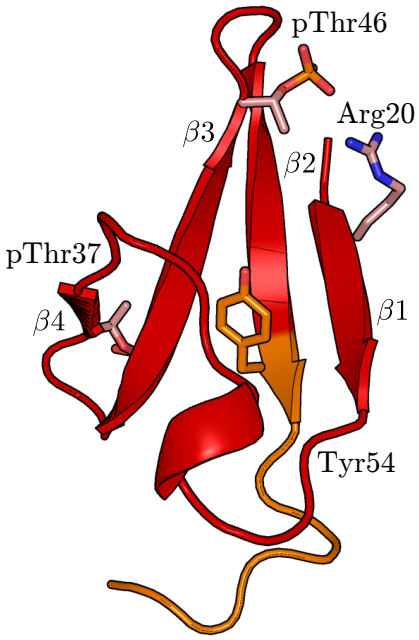


Figure 5.1.: MD equilibrated structure of the 4E-BP2 protein (red cartoon). The two phosphorylated threonines Thr37 and Thr46 (stick model) stabilize their respective turn via a network of hydrogen bonds. Interaction of Arg20 with pThr46 in simulations holds the sheet $\beta 1$ in place and prevents exposure of the eIF4E binding motif starting at Tyr54 (orange).

sequence motifs. When phosphorylated, hydrogen bonds of the phosphate with the rest of the motif impose a tight turn structure of the motif that is located at turns linking 3 of the 4 β-sheets.

In this work we investigated the stability of the folded structure of phosphorylated 4E-BP2 with molecular dynamics (MD) simulations. First, comparative replica-exchange molecular dynamics (H-REMD) simulations were employed to study the influence of phosphorylation on the stability of an isolated turn motif. The simulations indicate stabilization of the folded motif due to favorable interactions of the phosphate group with several other residues of the motif. The stability of the entire folded domain in the native un-phosporylated and phosphorylated form and with different mutations is investigated with multiple 200 ns simulations. While the phosphorylated 4E-BP2 remains folded in all simulations the wild type and phosphomimetic T37ET46E variant unfold in several simulations and allow us to characterize the de-phosphorylation-

induced unfolding process. The simulations also indicate an important role of three arginine residues (Arg20 and, to a lesser extend, Arg51, Arg56) that contact the phosphate groups in the folded state. An Arg20Ala substitution results in an increased tendency of unfolding and indicates a vital importance of the additional hydrogen bond between Arg20 and the phosphate group of pThr46 for stability. Finally, estimates of the electrostatic energy change upon phosphorylation show, that the folded state is energetically more beneficial for 4E-BP2 when phosphorylated. The simulation study reveals the important role of contacts at the surface of a protein to stabilize the folded structure and how this is utilized to mediate phosphorylation dependent conformational switches in the 4E-BP2 system.

5.2. Methods

Simulations were started from the experimental, phosphorylated pT37pT46 NMR structure with PDB code 2MX4[299] (Figure 5.1). Sequence variants with non-phosphorylated T37 and T46 (T37T46), an Arg20Ala (R20A) and a T37E/T46E substitution were generated in silico using PyMOL[256]. Each system was solvated in TIP4P-D[304] water, explicitly parametrized for intrinsically disordered proteins, and neutralized with Na+ and Cl- ions. MD simulations used a step size of 2 fs and the Particle Mesh Ewald method[136] for long range electrostatic interactions. In the equilibration procedure first energy minimization was performed for 5000 steps with the steepest descent algorithm. Next, 50000 steps of NVT equilibration were run with the Berendsen thermostat[120] at 300 K(turn motif) or 325 K(entire protein domain), where all heavy atoms were position-restrained. Finally, NPT equilibration with the Berendsen barostat[120] at 1.01 bar was run for 500000 steps with only the backbone atoms restrained.

5.2.1. Free Energy Calculations on Peptide Turn Unfolding

The free energy calculations were run with v4.6 of the GROMACS[305] package, using the amber99sb-ildn-star[212,250] force field for the protein with a cut-off of 11 Å. Free energy simulations were performed on the turn motif formed

by residues 44 to 53 of the E4-BP2 protein. Start and reference structure corresponded to the coordinates in the 2MX4 pdb structure (with NH3+ and COO- termini). We simulated three variants of the turn motif: Residue 46 (Thr) was either phosphorylated (pT46), non-phosphorylated (T46) or replaced by a phosphomimetic Glu residue (T46E substitution). All initial, folded structures were solvated with $\sim$ 2850 water molecules. The distance between C_α of end residues 44 and 53 as the reaction coordinate was enforced in 21 harmonic umbrella windows with a force constant of 500 kJ/mol between 5 - 25 Å, placed at distances 0.50 0.62 0.74 0.80 0.86 0.92 0.98 1.04 1.08 1.14 1.20 1.26 1.32 1.42 1.54 1.66 1.82 2.00 2.16 2.34 2.50Å. To enhance sampling, we used Hamiltonian replica exchange molecular dynamics (H-REMD) and attempted exchanges between umbrella windows every 500 steps (H-REUS method). Production runs were run for 50 ns per window, discarding the first 10 ns as equilibration. The potential of mean force (PMF) was calculated using the WHAM[306] algorithm.

5.2.2. Unrestraint Molecular Dynamics Simulations

Simulations of the entire folded 4E-BP2 domain were run with pmemd.cuda[307] as implemented in Amber14[308]. All variants of the 4E-BP2 domain, parametrized with the ff14SB[308] force field, were solvated in $\sim$ 38000 water molecules. Hydrogen mass repartitioning[119] in combination with SHAKE[118] constraints for all solute bonds without hydrogens allowed for a time step of 4 fs in production runs at a temperature of 325 K and pressure of 1.01 bar controlled with Berendsen baro- and thermostats. 10 simulations of 200 ns were run for all four sequence variants. Trajectories were visualized and analyzed with VMD[309].

5.2.3. Finite Difference Poisson Boltzmann Calculations

Effects of electrostatic interactions were estimated for the intra-molecular and solvent interactions after removal of explicit solvent. Intra-molecular electrostatic energies were calculated summing the pairwise electrostatic potential between all non-excluded atoms as implemented in cpptraj[310]. Solvent interactions were estimated by solving the finite-difference Poisson-Boltzmann (FDPB) with the adaptive Poisson-Boltzmann Solver[311] (APBS) and calculating the

difference of energies of the solvated and de-solvated molecule. Since FDPB calculations are sensitive to the placement of molecules in the finite difference grid we performed all calculations on the same ensemble of conformations and produced all sequence variants by eliminating/modifying only the corresponding side chains in the basic ensemble. This results in cancellation of errors due to different placements of conformations in the finite difference method. As the dielectric constant of the protein we used $\epsilon_p = 20$ to cover reorganization of the protein polar groups and water penetration[312]. Energies were calculated for one unfolded structure and an ensemble of 200 folded structures. The unfolded structure was taken from a snapshot of wild type 4E-BP2 where the protein was mostly and extended chain. The 200 folded structures were snapshots from the simulations of phosphorylated 4E-BP2.

5.3. Results

Phosphorylation can influence the stability of local structural motifs in proteins (formed by short sequence segments) such as alpha-helices or turn motifs. However, due to the negative charge of the phosphate group a long range electrostatic stabilization of the whole global protein fold is also possible. By using H-REMD Umbrella Sampling (H-REUS) free energy simulations we first investigated the influence of phosphorylation on the stability of a turn motif in 4E-BP2 followed by studying the stability of the whole domain by extended MD simulations.

5.3.1. Effects of Phosphorylation on a Turn Motif

The formation of tight turn motifs in the two identical pTPGGT sequences at T37 and T46 seems to be a key factor for the stability of the folded 4E-BP2 upon phosphorylation[299]. Free energy H-REUS simulations along a distance coordinate between the terminal amino acids were employed for turn opening/closing and to assess the influence of the phosphorylation on turn stability, using the peptide segment formed by residues 44-53. The H-REUS method along the reaction coordinate was used to improve convergence and effectively

allow folding and unfolding of the turn motif upon replica exchanges.

The calculated potential of mean force (PMF) or free energy along the reaction coordinate indicates a significant dependence on the modification of residue 46 (Figure 5.2). For the T46 and the pT46 variants conformations close to the native turn structure are the most stable conformations with the lowest free energy whereas for the T46E substitution unfolded conformations with large distance between terminal residues are of lower free energy than conformations close to the native structure. The minimum at $d \approx 6\,\text{Å}$ is defined by a H-bond between the backbone oxygen of residue 52 and the NH3 cap at Ser44 and differs slightly from the native turn structure. The second, global minimum at $d \approx 8\,\text{Å}$ is closest to the the turn motif from the NMR structure with three backbone H-bonds (44/O-52/H, 50/O-46/H, 46/O-49/H) and, for pThr46, additional H-bonds of the phosphate group. The third minimum at $d \approx 11\,\text{Å}$ represents an overall still collapsed state with the turn partially present but no stable backbone H-bonds. The non-phosphorylated turn structure (wt) is barely stable. Integrating the Boltzmann probability for folded and unfolded regimes, respectively, actually gives an overall positive folding free energy, favoring the unfolding of the isolated turn. The calculated free energy of transitions from the unfolded to the folded native like state is $2\,\text{kJ/mol}$, if states up to $d = 9\,\text{Å}$ are considered folded. Surprisingly the substitution T46E further destabilizes the turn motif and increases the folding free energy to $7\,\text{kJ/mol}$. In contrast, phosphorylation stabilizes the folded turn motif and also creates a significant energetic barrier towards opening. Turn folding is, in this case, energetically beneficial with a folding free energy of $-2\,\text{kJ/mol}$ upon phosphorylation.

Aside from formed H-bonds we took a closer look at replica 4 representing the folded state with its umbrella minimum at $d = 8.0\,\text{Å}$. In this replica, most of the time the methyl groups of pThr46 or Thr46 form a hydrophobic closed contact with the methyl group of Thr50 and the apolar side chain of Ile52. Glu46, lacking such a methyl group, cannot engage as well in such a hydrophobic interaction and is solvated more frequently. The root mean square fluctuations (RMSF), particularly of the apolar groups of Thr50 and Ile52, are increased for T46E compared to wt and pT, shown in Figure 5.3, pointing toward more flexibility and less stability of hydrophobic interactions.

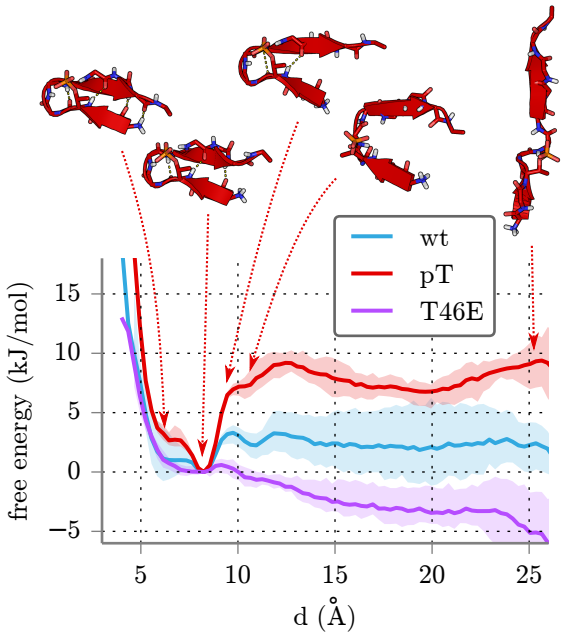


Figure 5.2.: Calculated free energy change (Potential of mean force: PMF) of the opening of a turn motif for the wild type, phosphorylated pT46 and the phosphomimetic mutation T46E. The reaction coordinate d was chosen as the distance between C_α-atoms of residues 44 and 53. Structures (from left to right) show a folded sub-state with an additional H-bond to the artificial N-terminal cap, the folded turn motif, the opening of the furthest β-sheet H-bond, a collapsed state without defined structure and a completely open turn loop. Secondary structure characterization was identical for all structures for clarity, i.e. the last three structures are not β-sheet like any more. Backbone H-bonds are shown in yellow dashes.

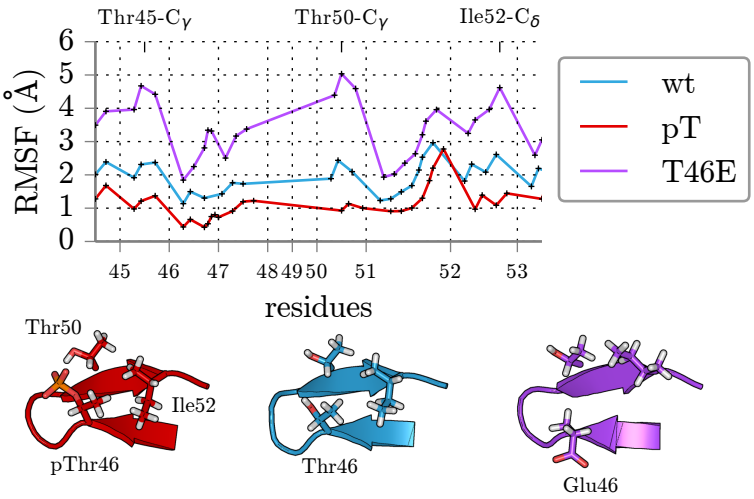


Figure 5.3.: Root mean square fluctuations (RMSF) of all heavy side chain atoms of the three simulated turn motifs. RMSF is calculated for the 4th replica where the umbrella minimum $d = 8\,\text{Å}$ matches the global free energy minumum and thus the stable turn motif. Values on the x-axis indicate residue numbers for orientation, but the RMSF was calculated for each side chain atom. Structures show snapshots with hydrophobic contacts that fluctuate more with the T46E mutation.

All side chains forming H-bonds are also present in the motif containing the other phosphorylation site Thr37 and the hydrophobic Ile52 finds an equivalent in Phe43, so we expect the free energy landscape to be similar.

5.3.2. Global Stability of Variants of Folded 4E-BP2

In order to compare the global stability of the folded 4E-BP2 domain we performed multiple unrestrained MD simulations of 200 ns at elevated temperature (325 K) for the domain variants. This allowed for qualitative insight into the stabilization due to phosphorylation. Simulations were performed on the non-phosporylated domain (wt), the pT37 and pT46 phosphorylated domain (pT), a variant with an Arg (residue 20) replaced by Ala (pT-R20A) and one variant with T37E and T46E substitutions (T37ET46E). The backbone root mean square deviations (RMSDs) with respect to the experimental structure[299] served as indicators for unfolding and are plotted in Figure 5.4. All simulations

of phosphorylated 4E-BP2 (pT) stay close to the native folded structure with an RMSD below 5 Å. In contrast, several of the wild type (wt) simulations show an unfolding of the protein and large RMSD fluctuations. Using an unfolding criterion of an RMSD from the native structure above 10 Å, in all pT simulations 4E-BP2 remained stable, but in 4 out of 10 wt simulations the domain unfolded. For a lower unfolding criterion of 5 Å RMSD 9/10 of the wt simulations unfold compared to 3/10 for the pT simulations. Independently of an explicit RMSD threshold the evolution of the average of the RMSD over all simulations in the lowest panel of Figure 5.4 shows that only the phosphorylated 4E-BP2 remains stable (compare red line with other lines in lowest panel of Figure 5.4). Remarkably, in all unfolding simulations of the wild type, unfolding starts with the detachment of the β1 strand and subsequent exposure of the binding motif. The two turn motifs at some point deform/unfold in all wt simulations, even if the overall fold of the protein is roughly conserved.

A means of investigating the structural impact of phosphorylation of Thr37 and Thr46 are the hydrogen bonds formed by the two amino acids with and without phosphorylation. Figure 5.5 reports all hydrogen bonds (within a distance cut-off of 3.0 Å) formed by (p)Thr37 and (p)Thr46, respectively. The hydrogen bonding pattern with other loop residues is similar for both phosphorylation sites, as expected since the sequence and the structural motif of the loops are identical. For the wild type, no significant binding to Gly39 is observed. The occasional H-bond to Gly40/49 is made by the backbone oxygen of the (phosphorylated) threonine and the backbone nitrogen of the glycine. H-bonds to Thr41/Thr50 are mostly formed as the β-sheet backbone H-bond, but also by side chain oxygens of both threonines. H-bonds for Thr46 are observed less compared to Thr37 as the loop containing Thr46 unfolded more easily. Apparently, adding the phosphate group with three free oxygens increases the number of possible hydrogen bonds significantly. The backbone nitrogen of Gly39/Gly48 forms a H-bond with the phosphate group in half of the simulation times. In addition to the backbone H-bond also seen without phosphorylation, Gly40/Gly49's nitrogen occasionally binds to an oxygen of the phosphate moiety. Contacts to Thr41/Thr50 strongly increase compared to the wild type as the phosphate moiety forms a stable hydrogen bond to the

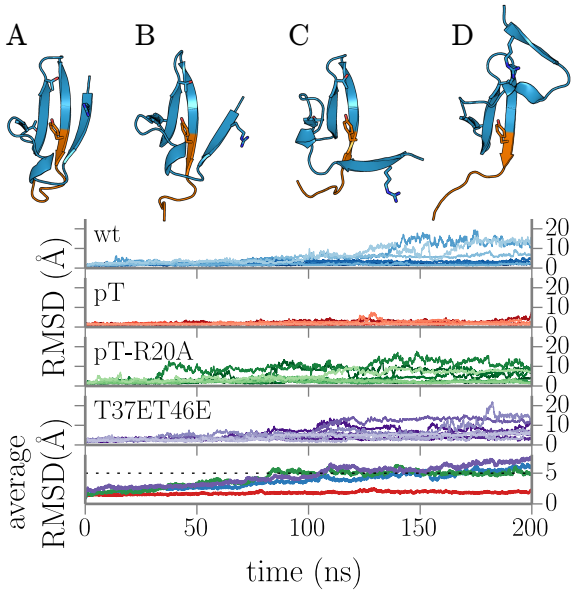


Figure 5.4.: Backbone RMSD of 10 simulations at $T = 325\,\mathrm{K}$ of 4E-BP2 for the wild type, the pT37pT46-phosphorylated protein, a phosphorylated variant where Arg20 is mutated to Ala and another variant where phosphorylated side chains are substituted by Glu residues. The bottom panel shows the average RMSD of all 10 simulations. Cartoon structures indicate characteristic snapshots from the unfolding pathway of the simulations: From the folded structure (A) $\beta 1$ unzips from the N-terminus (B) and completely detaches (C), followed by an opening of the bar covering the binding motif (orange) (D).

side chain oxygen of the threonine. Finally, pThr46 features another contact to Arg20 from the neighboring β-sheet β1. For pThr37 no tight H-bond outside of the turn motif were observed, but Arg51 and Arg56 are mostly pointing towards the phosphorylation site, resulting in electrostatic stabilisation. As an indication of an effect on these arginines the root mean square fluctuations (RMSF) of the carbon atom of side chains of Arg20, Arg51 and Arg56 are reduced in simulations of the phosphorylated protein relative to wild type, shown in Figure 5.6.

In the pT simulations Arg20 was identified to form stable H-bonds with pT46. In order to test the contribution of Arg20 to the stability of the 4E-BP2 domain we also performed 10 independent 200 ns simulations on the phosphorylated 4E-BP2 with an R20A substitution. Unfolding was observed in 5 out of 10 simulations(Figure 5.4) with the unfolding criterion of backbone RMS deviation above 10 Å from the native fold (9/10 for RMSD 5 Å threshold). The average RMSD resembles the evolution of wild type 4E-BP2. Hydrogen bonding of the phosphorylated threonines shows identical H-bonds are formed apart from the broken Arg20-pThr46 connection and H-bonds were equally stable over the simulation time. Unfolding in each MD run did occur via a detachment of β1 from the parallel β4, as observed for the wild type. Hence, the simulations indicate that Arg20 plays a key role for stabilizing the 4E-BP2 fold by preventing rapid dissociation of the N-terminal β1 strand in the early stage of unfolding.

Substitution of phosphorylated threonines with glutamic or aspartic acid as phosphomimetics did not show any sign of folding in the experimental NMR study of 4E-BP2[299]. Indeed, our free energy calculations of the isolated turn motif sequence show, that the turn motif with phosphomimetic Glu is even less stable than the wild type (see previous paragraph). The turn conformation is not the global free energy minimum as Glu46 rarely forms stabilizing H-Bonds with Thr50 but instead even disfavors the turn state. Furthermore, multiple 200 ns MD simulations at elevated temperature on the 4E-BP2 domain variant with the T37E and T46E substitutions indicate similar behavior as the wild type. The respective backbone RMSD and hydrogen bonding pattern are shown in Figures 5.4 and 5.5. In 3/10 cases the protein unfolded to a RMSD larger

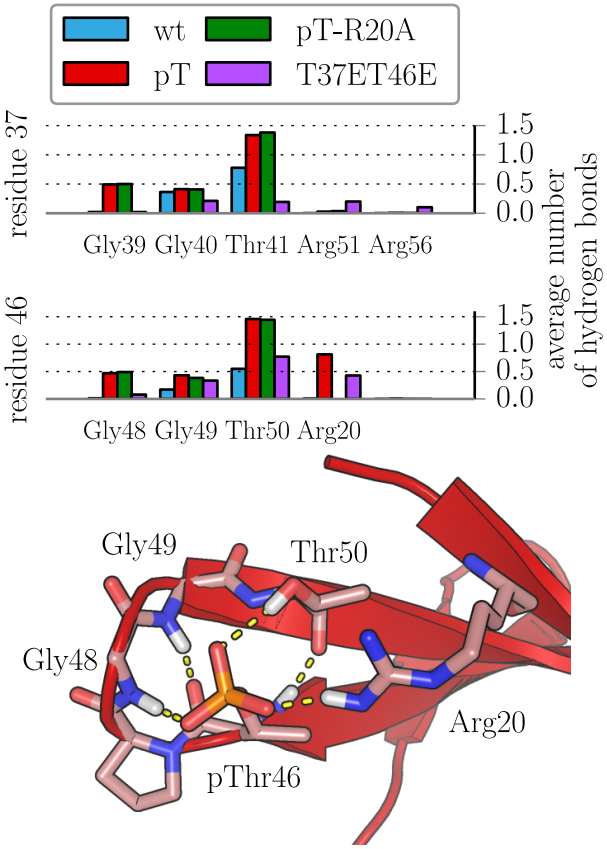


Figure 5.5.: All H-bonds formed by the two phosphorylation sites Thr37 and Thr46 during unrestraint MD simulations. Both show identical patterns for the loop stabilizing contacts but pThr46 additionally features a hydrogen bond with Arg20 from the neighboring β1-strand. Note, that this interaction stabilizes the whole β-sheet. The cartoon structure shows the folded turn motif around pThr46. Only hydrogen atoms relevant for H-bonds are depicted.

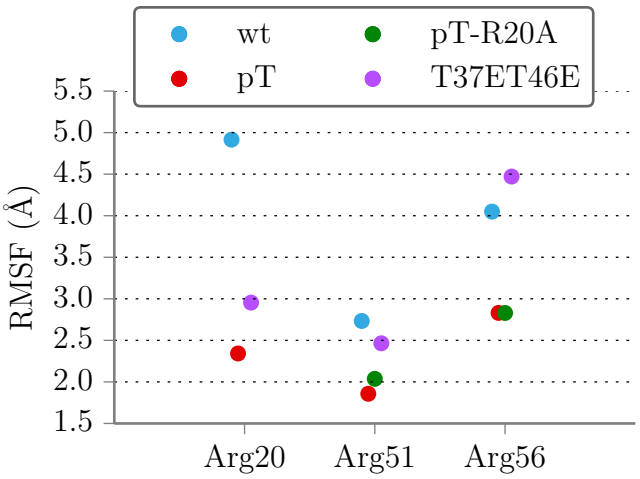


Figure 5.6.: Root mean square fluctuations of CZ of three arginines interacting with the phosphorylation site. To avoid artifacts from already unfolded conformations for all four variants of 4E-BP2 only simulations remaining stable were considered.

than 10 Å (9/10 for RMSD 5 Å). Hydrogen bonding patterns differ for the two phosphorylation sites, as opposed to the other three systems. Glu46 produced a similar hydrogen bonding pattern as pThr46, contacting turn residues as well as Arg20, albeit with significantly lower occupancy, as typically only one of the bonds to Arg20 or the turn motif was formed. Glu37, however, formed less hydrogen bonds to the other turn residues than the backbone hydrogen bonds of wild type and instead contacted the spatially more distant, positively charged Arg51 and Arg56. These two arginines also pointed in the direction of pThr37 in the pT simulations, but do not reach close enough to form hydrogen bonds. Glu37 is less tightly bound to the turn motif than pThr37 and the negative charge can reach closer to the arginines to form actual H-bonds. The distance of Glu37 to both arginines was on average, however, larger than that of pThr37 in the respective simulations and the arginine side chains are more flexible compared to phosphorylated Thr37/46 residues (see SI Figure 5.6).

5.3.3. Electrostatic Contributions

Phosphorylation significantly alters the charge distribution of the 4E-BP2 domain. To assess electrostatic effects we estimated the electrostatic energy difference between the unfolded state and an ensemble of folded states taken from the pT MD simulations. Electrostatic reaction field contributions due to interactions with the solvent were calculated by solving the finite-difference Poisson-Boltzmann (FDPB) equation using the adaptive Poisson-Boltzmann Solver APBS[311] (see methods for details). The intra-molecular electrostatic Coulomb energies of the protein were determined from pair-wise electrostatic energies between all non-bonded atoms. The same structures for the unfolded reference and the folded ensemble were used for all cases. Errors are calculated as the standard deviation of energies obtained for the folded ensemble.

The calculated electrostatic energy differences between folded and unfolded states indicates that electrostatic interactions overall disfavor folding (Table 5.1). For the wild type this is due to both increased Coulomb energies and increased reaction field energies. Upon phosphorylations folding is favored in terms of improved Coulomb interaction energy, but still penalized by the less favorable reaction field of the solvent, as charges are more exposed to solvent in the unfolded state. Although the results suggest that, overall, electrostatics disfavor folded structures, double phosphorylation at T37 and T46 reduces the energetic penalty by almost 20 kJ/mol. Single phosphorylation at either T37 or T46 resulted in electrostatic interactions of similar magnitude as the wild type. This is in line with experimental results indicating that only double phosphorylation leads to a significant stabilization of the folded domain. The phosphomimetic T37ET46E mutations slightly reduced the overall electrostatic folding penalty compared to the wild type, but are not as effective as phosphorylations. The R20A mutation on pT shows similar penalization by the reaction field but drastically reduced coulombic benefits of folding and has the highest electrostatic folding costs of all systems.

Visualizations of the electrostatic potential on the surface of final folded conformations of all four simulated systems, obtained from APBS calculations, are shown in Figure 5.7. For the wild type both phosphorylation sites as well as the β1 strand have a positive surface charge. When phosphorylated, both

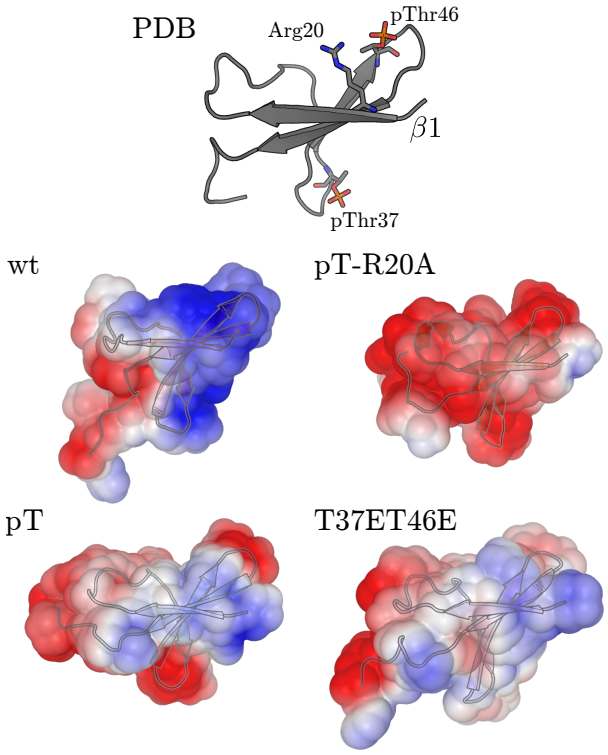


Figure 5.7.: Surface charges for all simulated systems. The top cartoon shows the experimental structure in similar orientation for reference. Surface potentials are obtained from approximate solutions of the finite difference Poisson-Boltzmann equation and are colored from -3 (red) to +3 (blue) KBT potential units.

turns containing the phosphorylation sites are negatively charged and the Thr46-loop attracts β1. The R20A mutation de-charges the surface of β1 almost completely. The phosphomimetic T37ET46E mutations almost neutralize the two turn motifs and lead to a neutral protein surface compared to the wild type or phosphorylated 4E-BP2. The C-terminal binding motif (bottom left in Figure 5.7) is very flexible, slightly more so in the non-phosphorylated simulations. A specific affinity to remain in tighter positions, that could be inferred from Figure 5.7, could not be observed.

system	$dE_{\text{fold-unfold}}$ (kJ/mol) dE_{rf}	dE_{coul}	dE
wt	10	20	29 ± 12
pT	70	-62	9 ± 10
pT37	40	-16	24 ± 9
pT46	42	-5	38 ± 14
T37ET46E	39	-17	21 ± 11
R20A	72	-30	42 ± 10

Table 5.1.: Electrostatic energy of folding

5.4. Discussion

Post-translational modifications have been shown to affect the secondary structural elements in intrinsically disordered proteins[107]. The 4E-BP2 domain, intrinsically disordered in the wild type non-phosphorylated form, has been found to not only change properties of a secondary structure element but to fold completely upon phosphorylation[299]. Here we applied several simulation methods to investigate effects of the phosphorylations in atomic detail.

As suggested by Bah et al.[299] the phosphorylation of two threonines stabilizes two similar turn motifs allows for several hydrogen bonds to be formed with the phosphate moiety. Both phosphorylation sites are followed by the same PGGT sequence and create comparable hydrogen bonds. Bah et al.[299] roughly estimated the free energy of unfolding the phosphorylated protein to be in the range of 0 - 2 kJ/mol. Our calculated free energy of 4 kJ/mol for unfolding the phosphorylated turn motif compared to the wild type agrees with this estimate, as the energy obtained for the segment would be reduced by a loss of entropy of the full protein. The agreement of the folding free energies indicates, that the turn motif is formed without further folding of the protein if only one threonine is phosphorylated.

However, effects of the phosphorylation do not end at the stabilization of the turn motifs linking β2 with β3 and β3 with β4. Additionally, the phosphate group of pThr46 holds β1 in its parallel sheet with β4 through non-bonded and non-local interactions with Arg20. Residues Arg51 and Arg56 perform a similar function in case of the pT37 residue but are not precisely in H-bonding distance during the simulations. However, the strong electrostatic

interaction was reflected in reduced flexibility in the pT vs. wt simulations. The importance of these residues for structure formation and thus eIF4E binding and translation inhibition is underlined further by the fact that all three Arg residues interacting with the phosphorylation sites are conserved across multiple species[313]. Furthermore, phosphorylation drastically alters the charge distribution in the protein and the altered electrostatic interactions alleviate folding for phosphorylated 4E-BP2.

The lack of a third H-Bond partner for Arg20 can explain why substitution of the phosphorylation sites by phosphomimetics, T37ET46E or T37DT46D, did not show any signs of folding in experiment[299]. Indeed simulations with a T37ET46E mutation showed that occasionally formed hydrogen bonds are not stable enough to maintain the folded turn motif. Furthermore, the phosphomimetics did not simultaneously establish bonds to the turn motif and the stabilizing arginines Arg20 or Arg51/Arg56.

While we are confident that equilibrium is reached with the advanced sampling techniques applied to the isolated loops, we do not expect the 10 continuous MD simulations per system to represent the equilibrium. For the phosphorylated 4E-BP2, Bah and colleagues[299] found indications of an exchange between major and minor conformers, so the phosphorylated protein might also be unfolded at times. One of our trajectories of phosphorylated 4E-BP2 did show signs of detaching β1. The overall higher stability of phosphorylated 4E-BP2, however, can clearly be seen from our simulations.

5.5. Conclusion

Our simulations explain the mechanism of fold stabilization by phosphorylation of 4E-BP2 due to specific local effects on two tight turn motif connecting the β-strands and a global electrostatic stabilization of the entire folded form. The simulation results are in good agreement with the experimental observation that only a double-phosphorylation of 4E-BP2 effectively stabilizes the folded form. The simulations on the T37ET46E variants indicate that in the case of 4E-BP2 these substitutions abrogated the stabilizing benefits of the phosphorylations not acting as effective mimetics because of a reduced ability

to form stable H-bonds and less close van der Waals contacts. It also explains the experimental observation that this variant does not fold. Our simulations suggest further that the Arg residues in 4E-BP2, although located at the surface, are critical for stable folding. Specifically, our prediction that the R20A substitution strongly destabilizes the protein could be tested experimentally. The combined local and global stabilizing effects of specific phosphorylation might be characteristic for other phosphorylation dependent structural motifs in proteins. These principles could useful for identification of structural motifs in proteins that can be stabilized by phosphorylation and for the design of motifs that undergo phosphorylation induced conformational changes.

6. Summary and Outlook

Intrinsically disordered proteins are a recent and fascinating topic in modern biophysics. In this work we attempted to study some representatives of IDPs with both traditional and new methods of molecular dynamics simulations.

The empirical and classical force fields employed to model proteins in their aqueous environment are able to reproduce the intrinsic disorder, but largely differ in the amount of helical content. In our simulations of 11 segments of the IDP Axin the common TIP3P water model did predict the amount of helicity in good agreement with NMR measurements. Newly developed water force fields and especially TIP4P-D did sample more open and extended structures, but come at a price of correct helical content in the short peptides. Only with adaptions to the backbone torsional potentials, as done for the amber99ws force field, the higher helical content can be reproduced.

For a system of coupled folding and binding, the pKID-KIX complex, our simulation data suggests that phosphorylation does shift the conformational equilibrium of pKID towards the bound state in the immediate vicinity of the phosphorylation site. This indicates that not all structure formation in pKID must happen on the surface of KIX and that the binding process might involve some conformational selection. A simulation of the association process, however, also showed that structure formation after an initial key contact is possible. So the association of pKID and KIX is in part a conformational selection through phosphorylation and an also a case of induced fit.

As another example of the effects of phosphorylation the stability of folded 4E-BP2 was investigated. Our advanced sampling simulations confirmed that the two phosphorylations stabilize similar turn motifs. Furthermore, in free simulations we identified interactions with other surface arginine side chains that are crucial for the stability of the fold. Electrostatic calculations supported

the view of a stabilizing effect of the additional phosphate groups. Substitution of the phosphorylations with aspartic acid did not recover the stabilizing effects, underlining the unique properties of phosphate groups.

As became clear during the investigations of this work the simulation of IDPs still faces some problems. Force fields are not yet accurate enough and optimizations with respect to one property of IDPs, e.g. the radius of gyration, might worsen the representation of other properties. The diverse nature of IDPs, ranging from compact molten globules to extended random coils, will continue to pose a problem in the search of suitable force field parameters. The difficulties of obtaining precise experimental data that can be compared with simulations pose a further problem for force field optimization. Even with correct force fields the time scales accessible to current simulations are not nearly enough to cover the conformational dynamics of IDPs and the exchange between conformational sub-states. Already our microsecond simulations of free pKID show that conformations of relatively small systems can be stable for hundreds of nanoseconds. Many conversions between substates must to be sampled for statistical validity, so converged simulations are a distant prospect for IDPs and more powerful techniques of advanced sampling will have to be applied.

Nevertheless, simulations consistently and correctly predicted disorder where it was found in experiments. While the complete conformational ensemble was inaccessible, the stability of specific conformations could successfully be compared for different phosphorylation states.

For the future MD simulations are on their way to becoming a reliable and valuable tool for biophysicists to learn about IDPs. The force field problem is being tackled by several groups around the world and better force fields are currently developed[314]. Since there are no fundamental differences to folded proteins, which today can be described quite well with the approximations of MD, an adaption to IDPs should be feasible. The future is looking even better on the sampling front. Computer power has steadily increased over the last decades, and with the transition to calculations on GPUs the highly parallelizable MD simulations will further benefit from the rapid development in this area.

A. Implementation of the dRMSD

The dRMSD as a reaction coordinate for umbrella sampling with replica exchange has been implemented in GROMACS 4.6.2[251].

Equations of the dRMSD potential

As stated in Chapter 2, Equation 2.34, the distance root mean square deviation (dRMSD) is defined as

$$R(d_1, \ldots, d_N) = \sqrt{\frac{1}{N}\sum_{i}^{N}(d_i - d_{i0})^2} \tag{A.1}$$

Where index i runs over all distances between the N atom pairs that contribute to the dRMSD.

The harmonic potentials along R are of the form

$$V(d_1, \ldots, d_N) = \frac{1}{2}k_0\big(R(d_1, \ldots, d_N) - R_0\big)^2 \tag{A.2}$$

with a specific reference dRMSD R_0 and the distance d_i being a function of the coordinates $\mathbf{r}_{i1}, \mathbf{r}_{i2}$ of the two atoms of the pair:

$$d_i \equiv d_i(r_{i1}, r_{i2}) = |\mathbf{r}_{i1} - \mathbf{r}_{i2}| = \sqrt{(\mathbf{r}_{i1} - \mathbf{r}_{i2})^2} \tag{A.3}$$

The forces on atom coordinate x of atom $i1$ is then calculated from

$$\begin{aligned} F_{x_{i1}} &= -\frac{dV(d_1,\ldots,d_N)}{dx_{i1}} \\ &= -\frac{k_0}{N}\cdot\frac{R(d_1,\ldots,d_N)-R_0}{R(d_1,\ldots,d_N)}\cdot(d_i-d_{i0})\cdot\frac{x_{i1}-x_{i2}}{d_i} \\ &= -F_{x_{i2}} \end{aligned}$$

Thus the vectorial force is given by

$$\mathbf{F}(d_i) = -\frac{k_0}{N}\cdot\frac{R(d_i)-R_0}{R(d_i)}\cdot(d_i-d_{i0})\cdot\frac{\mathbf{d}_i}{d_i} \tag{A.4}$$

Finally, the contribution to the Hamiltonian of this distance RMSD potential is

$$\mathcal{H}_{\text{dRMSD}} = V(d_1,\ldots,d_N) \tag{A.5}$$

Lambda Scaling Along the dRMSD

For the application of US typically several windows along the reaction coordinate are defined via a transition coordinate λ. We defined the λ-dependence of R as

$$R(d_i,\lambda) = \sqrt{\frac{1}{N}\sum_i^N\left(d_i-(1-\lambda)d_{i0}^A-\lambda d_{i0}^B\right)^2} \tag{A.6}$$

The λ-dependent distance RMSD potential has then the form

$$V(d_i,\lambda) = \frac{k_0}{2}\left(R(d_i,\lambda)-(1-\lambda)R_0^A-\lambda R_0^B\right)^2 \tag{A.7}$$

This allows transitions from one state with reference distances d_{i0}^A to another state with reference distances d_{i0}^B. Also, with no d_{i0}^B defined, a continuous sampling of λ in the range $[0,1]$ allows sampling from the structure defined with distances d_{i0}^A to unfolded structures up to a dRMSD deviation of R_0^A.

The derivative of the potential with respect to λ then is

$$\begin{aligned}\frac{dV(d_i,\lambda)}{d\lambda} =& k_0\left(R(d_i,\lambda)-(1-\lambda)R_0^A-\lambda R_0^B\right)\cdot \\ & \left(\frac{dR(d_i,\lambda)}{d\lambda}+R_0^A-R_0^B\right) \\ =& k_0\left(R(d_i,\lambda)-(1-\lambda)R_0^A-\lambda R_0^B\right)\cdot \\ & \left(\frac{1}{2NR(d_i,\lambda)}\left(2\sum_i^N\left(d_i-(1-\lambda)\,d_{i0}^A-\lambda d_{i0}^B\right)\left(d_{i0}^A-d_{i0}^B\right)\right)+\right. \\ & \left. R_0^A-R_0^B\right)\end{aligned} \tag{A.8}$$

and the force in direction x for atom $i1$ is

$$\begin{aligned}F_{x_{i1}} =& -\frac{dV(d_i,\lambda)}{dx_{i1}} \\ =& -\frac{k_0}{N}\cdot\frac{R(d_i,\lambda)-(1-\lambda)R_0^A-\lambda R_0^B}{R(d_i,\lambda)}\cdot \\ & \left(d_i-(1-\lambda)d_{i0}^A-\lambda d_{i0}^B\right)\cdot\frac{x_{i1}-x_{i2}}{d_i} \\ =& -F(x_{i2})\end{aligned} \tag{A.9}$$

and thus the vectorial force is

$$\begin{aligned}\mathbf{F}(d_i,\lambda) =& -\frac{k_0}{N}\cdot\frac{R(d_i,\lambda)-(1-\lambda)R_0^A-\lambda R_0^B}{R(d_i,\lambda)}\cdot \\ & \left(d_i-(1-\lambda)d_{i0}^A-\lambda d_{i0}^B\right)\cdot\frac{\mathbf{d}_i}{d_i}\end{aligned} \tag{A.10}$$

Gromacs Parameters

To run a simulation with the adapted GROMACS code the following settings are required.

MDP Parameters

`drmsd-pot: yes/no`
Enable/Disable distance based RMSD potential.

`drmsd-ref: (0.0) [nm]`
Reference distance RMSD for state A

`drmsd-refB: (0.0) [nm]`
Reference distance RMSD for state B

`drmsd-k0: (1000.0) [kJ mol`$^{-1}$ `nm`$^{-2}$`]`
Force konstant of the distance RMSD potential

`nstdrmsdpout: (100)`
Frequency of writing distance RMSD potential output

`bonded-lambdas = 0.00 0.20 0.40 0.60 0.80 1.00`
λ values for umbrella windows. (requires further free energy parameters)

Topology File

The topology needs to include a `[ drmsd_restraints ]` section:

```
Topology file parameters:
[ drmsd_restraints ]
;   ai      aj      type    d0      d0B (optional)
    9       39      1       0.5     0.6
    19      49      1       0.5     0.6
    29      59      1       0.5     0.6
```

Atom pair `ai`, `aj` which forms a distance pair. Default function type 1. `d0` is the reference distance of this atom pair. Note that `ai`, `aj` are molecule interal indices. The first atom of each molecule has index 1.

mdrun Parameters

The possible long distance bonds that have to be calculated each step clash with the domain decomposition principle of Gromacs. Simulations will not start without the explicit request of particle decomposition:

```
mdrun [...] -pd
```

g_drmsd

The implementation of the distance restraint can write out the dRMSD calculated during the simulation. Additionally the tool `g_drmsd` can be used to calculate dRMSDs from a given trajectory. To obtain the distances and applied forces use `g_drmsd`. `g_drmsd` has to be given at trajectory and a run input file with all the settings for the dRMSD method. The tool then for each frame of the trajectory extracts the dRMSD and the resulting potential to a output file.

```
-f Input, trajectory: .xtc, .trr etc.
-s Input, run input file: .tpr
-o Output file (drmsd.xvg), optional
```

If `g_drmsd` is given a list of trajectories and tpr files it will calculate the drmsd and potential for the first given trajectory with the first tpr and so forth. Non-matching numbers after the last underscore, e.g. traj_1.xtc and topol_2.tpr will give an error. Output for each trajectory will be written to files with matching number.

Acronyms

ABF adaptive biasing force.

AFM atomic force microscopy.

BAR Bennett acceptance ratio.

CD circular dichroism.

dRMSD root mean square deviation of a set of distances.

ER endoplasmatic reticulum.

FCS fluorescence energy transfer.

FEP free energy perturbation.

FRET Förster resonance energy transfer.

GPU graphics processing unit.

H-REMD Hamiltonian replica exchange molecular dynamics.

IDP intrinsically disordered protein.

IDR intrinsically disordered region.

MC Monte Carlo.

MD molecular dynamics.

MoRF molecular recognition features.

NMR nocleo-magnetic resonance.

NOE nuclear overhauser effect.

PME particle mesh Ewald.

PMF potential of mean force.

PRE paramagnetic relaxation enhancement.

PTM post-translational modification.

RDC residual dipolar couplings.

REMD replica exchange molecular dynamics.

REUS replica exchange umbrella sampling.

RMSD root mean square deviation.

RMSF root mean square fluctuations.

SAXS small angle X-ray scattering.

SLiM short linear motif.

T-REMD temperature replica exchange molecular dynamics.

TI thermodynamic integration.

US umbrella sampling.

List of Figures

List of Tables

List of Publications

Peer-Reviewed Publications

1 **R. Bomblies**, M. P. Luitz, and M. Zacharias. "Molecular Dynamics Analysis of 4E-BP2 Protein Fold Stabilization Induced by Phosphorylation." *Journal of Physical Chemistry B* (2016), in press.

2 **R. Bomblies**, M. P. Luitz, and M. Zacharias. "Mechanism of pKID/KIX Association Studied by Molecular Dynamics Free Energy Simulations." *Journal of Physical Chemistry B* 120.33 (Apr. 2016), pp. 8186–8192.

3 M. P. Luitz, **R. Bomblies**, E. Ramcke, A. Itzen, and M. Zacharias. "Adenylylation of Tyr77 Stabilizes Rab1b GTPase in an Active State: A Molecular Dynamics Simulation Analysis." *Scientific reports* 6 (Jan. 2016), p. 19896.

4 **R. Reisenauer**, K. Smith, and R. A. Blythe. "Stochastic Dynamics of Lexicon Learning in an Uncertain and Nonuniform World." *Physical Review Letters* 110.25 (2013), p. 258701.

Publications in Preparation

1 F. Zeller, M. P. Luitz, **R. Bomblies**, and M. Zacharias. "Accurate Multi-scale Simulation of Receptor-drug Association Kinetics: Application to Neuramidase Inhibitors." *submitted to Angewandte Chemie International Edition* (2016).

2 **R. Bomblies**, M. P. Luitz, S. Scanu, T. Madl, and M. Zacharias. "Transient Helicity in Intrinsically Disordered Axin-1 Depends on Force Field." *in preparation* (2016).

3 M. P. Luitz, A. Barth, A. H. Crevenna, **R. Bomblies**, D. Lamb, and M. Zacharias. "Covalent Dye Attachment Influences the Dynamics and Conformational Properties of Flexible Peptides." *in preparation* (2016).

4 M. P. Luitz, **R. Bomblies**, and M. Zacharias. "From Chaos to Order: The Association Process of RNAse-S Studied by Molecular Dynamics Simulations." *in preparation* (2017).

Book Chapters

1 C. N. Cavasotto, **R. Bomblies**, M. Luitz, and M. Zacharias. "Free Energy Calculations of Ligand–Protein Binding." In: *In Silico Drug Discovery and Design: Theory, Methods, Challenges, and Applications*. CRC Press, 2015, pp. 313–335.

Conferences

1 F. Zeller, **R. Bomblies**, M. P. Luitz, and M. Zacharias. "Influenza Neuraminidase Inhibitor Binding Studied by Molecular Dynamics Simulations." *SFB863 Workshop, Schloss Ringberg* (2015). Conference Talk.

2 **R. Bomblies**, M. P. Luitz, and M. Zacharias. "The Mechanism of pKID-KIX Complex Formation studied by Molecular Dynamics Simulations." *SFB1035 Annual Meeting*. Conformational Switches (July 2015). Garching, Germany, Poster Presentation.

3 Luitz, M. P., **R. Bomblies**, and M. Zacharias. "From Chaos to Order: The Association Process of RNAse-S studied by MD Simulations." *SFB1035 Annual Meeting*. Conformational Switches (July 2015). Garching, Germany, Poster Presentation.

4 **R. Bomblies**, M. P. Luitz, and M. Zacharias. "The Mechanism of pKID-KIX Complex Formation studied by Molecular Dynamics Simulations." *SFB1035 International Conference*. Conformational Switches (May 2015). Venice, Italy, Poster Presentation.

5 Luitz, M. P., **R. Bomblies**, and M. Zacharias. "From Chaos to Order: The Association Process of RNAse-S studied by MD Simulations." *SFB1035 International Conference*. Conformational Switches (May 2015). Venice, Italy, Poster Presentation.

6 **R. Bomblies**, M. P. Luitz, and M. Zacharias. "Characterization of Transiently Stable Structural Motifs in Intrinsically Disordered Proteins using Free Energy Simulations." *Biophysical Society Meeting* (Feb. 2015). Baltimore, USA, Poster Presentation.

7 Luitz, M. P., **R. Bomblies**, M. Zacharias, E. Bender, and A. Itzen. "Conformational Transitions in Switch Regions of the Ras-Like GTPase Rab1B Studied by Free Energy Simulations." *Biophysical Society Meeting* (Feb. 2015). Baltimore, USA, Conference Talk.

8 **R. Bomblies**, M. P. Luitz, and M. Zacharias. "The Mechanism of pKID-KIX Complex Formation Studied by Molecular Dynamics Simulations." *Biophysical Society Thematic Meeting*. Disordered Motifs and Domains in Cell Control (Oct. 2014). Dublin, Ireland, Poster Presentation.

9 Luitz, M. P., **R. Bomblies**, and M. Zacharias. "From Chaos to Order: The Association Pathway of RNAse-S." *Biophysical Society Thematic Meeting*. Disordered Motifs and Domains in Cell Control (Oct. 2014). Dublin, Ireland, Poster Presentation.

10 **R. Bomblies**, M. P. Luitz, and M. Zacharias. "Coupled Association and Structure Formation studied by Molecular Dynamics Simulations." *SFB1035 Annual Meeting*. Conformational Switches (July 2014). Garching, Germany, Poster Presentation.

11 Luitz, M. P., **R. Bomblies**, and M. Zacharias. "In Silico Investigation of Post-translational Tyr77 Modification on Rab1b Switching Mechanism." *SFB1035 Annual Meeting*. Conformational Switches (July 2014). Garching, Germany, Poster Presentation.

12 **R. Bomblies**, M. P. Luitz, and M. Zacharias. "MD Studies on Intrinsically Disordered Proteins." *SFB1035 PhD retreat*. Conformational Switches (Apr. 2014). Spitzingsee, Germany, Conference Talk.

13 Luitz, M. P., **R. Reisenauer**, and M. Zacharias. "Coupled Folding and Association: RNAse-S and pKID-KIX studied with MD Simulations." *Faltertage* (Oct. 2013). Regensburg, Germany, Poster Presentation.

14 Luitz, M. P., **R. Reisenauer**, and M. Zacharias. "Coupled Folding and Association: RNAse-S and pKID-KIX studied with MD Simulations." *SFB1035 Annual Meeting*. Conformational Switches (July 2013). Garching, Germany, Poster Presentation.

15 Luitz, M. P., **R. Reisenauer**, and M. Zacharias. "Association of Intrinsically Disordered Proteins: A Case Study of RNAse-S." *SFB1035 PhD retreat*. Conformational Switches (Mar. 2013). Aschau, Germany, Conference Talk.

16 **R. Reisenauer** and M. Zacharias. "The intrinsically disordered state of pKID." *Hünfeld Workshop*. Computer Simulation and Theory of Macromolecules (Apr. 2013). Hünfeld, Germany, Poster Presentation.

Bibliography

1 C. B. Anfinsen. "Principles That Govern the Folding of Protein Chains." *Science* 181 (1973), pp. 223–230.

2 R. W. Kriwacki, L. Hengst, L. Tennant, S. I. Reed, and P. E. Wright. "Structural Studies of p21Waf1/Cip1/Sdi1 in the Free and Cdk2-bound State: Conformational Disorder Mediates Binding Diversity." *Proceedings of the National Academy of Sciences of the United States of America* 93.21 (1996), pp. 11504–11509.

3 P. E. Wright and H. J. Dyson. "Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-function Paradigm." *Journal of Molecular Biology* 293.2 (1999), pp. 321–331.

4 V. N. Uversky, J. R. Gillespie, and A. L. Fink. "Why Are "Natively Unfolded" Proteins Unstructured Under Physiologic Conditions?" *Proteins: Structure, Function, and Bioinformatics* 41.3 (2000), pp. 415–427.

5 A. Dunker, J. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, and Z. Obradovic. "Intrinsically Disordered Protein." *Journal of Molecular Graphics and Modelling* 19.1 (2001), pp. 26–59.

6 P. Tompa. "Intrinsically Unstructured Proteins." *Trends in Biochemical Sciences* 27.10 (2002), pp. 527–533.

7 M. Sandal, F. Valle, I. Tessari, S. Mammi, E. Bergantino, F. Musiani, M. Brucale, L. Bubacco, and B. Samori. "Conformational Equilibria in Monomeric α-synuclein at the Single-molecule Level." *PLoS Biology* 6.1 (2008), e6.

8 V. Uversky. "Protein Folding Revisited. A Polypeptide Chain at the Folding–misfolding–nonfolding Cross-roads: Which Way to Go?" *Cellular and Molecular Life Sciences* 60.9 (2003), pp. 1852–1871.

9 S. L. Crick, M. Jayaraman, C. Frieden, R. Wetzel, and R. V. Pappu. "Fluorescence Correlation Spectroscopy Shows That Monomeric Polyglutamine Molecules Form Collapsed Structures in Aqueous Solutions." *Proceedings of the National Academy of Sciences of the United States of America* 103.45 (2006), pp. 16764–16769.

10 H. T. Tran, A. Mao, and R. V. Pappu. "Role of Backbone- Solvent Interactions in Determining Conformational Equilibria of Intrinsically Disordered Proteins." *Journal of the American Chemical Society* 130.23 (2008), pp. 7380–7392.

11 B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, and V. N. Uversky. "PONDR-FIT: A Meta-predictor of Intrinsically Disordered Amino Acids." *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1804.4 (2010), pp. 996–1010.

12 R. Pancsa and P. Tompa. "Structural Disorder in Eukaryotes." *PLOS ONE* 7.4 (2012), e34687.

13 P. Tompa. "Intrinsically Disordered Proteins: A 10-year Recap." *Trends in Biochemical Sciences* 37.12 (2012), pp. 509–516.

14 C. Bracken. "NMR Spin Relaxation Methods for Characterization of Disorder and Folding in Proteins." *Journal of Molecular Graphics and Modelling* 19.1 (2001), pp. 3–12.

15 D. Eliezer. "Biophysical Characterization of Intrinsically Disordered Proteins." *Current Opinion in Structural Biology* 19.1 (2009), pp. 23–30.

16 M. R. Jensen, M. Zweckstetter, J.-r. Huang, and M. Blackledge. "Exploring Free-energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy." *Chemical Reviews* 114.13 (2014), pp. 6632–6660.

17 M. R. Jensen, L. Salmon, G. Nodet, and M. Blackledge. "Defining Conformational Ensembles of Intrinsically Disordered and Partially Folded Proteins Directly From Chemical Shifts." *Journal of the American Chemical Society* 132.4 (2010), pp. 1270–1272.

18 J. A. Marsh and J. D. Forman-Kay. "Structure and Disorder in an Unfolded State Under Nondenaturing Conditions From Ensemble Models Consistent With a Large Number of Experimental Restraints." *Journal of Molecular Biology* 391.2 (2009), pp. 359–374.

19 M. D. Mukrasch, P. Markwick, J. Biernat, M. von Bergen, P. Bernadó, C. Griesinger, E. Mandelkow, M. Zweckstetter, and M. Blackledge. "Highly Populated Turn Conformations in Natively Unfolded Tau Protein Identified From Residual Dipolar Couplings and Molecular Simulation." *Journal of the American Chemical Society* 129.16 (2007), pp. 5235–5243.

20 F. N. Newby, A. De Simone, M. Yagi-Utsumi, X. Salvatella, C. M. Dobson, and M. Vendruscolo. "Structure-Free Validation of Residual Dipolar Coupling and Paramagnetic Relaxation Enhancement Measurements of Disordered Proteins." *Biochemistry* 54.46 (2015), pp. 6876–6886.

21 D. Ganguly and J. Chen. "Structural Interpretation of Paramagnetic Relaxation Enhancement-derived Distances for Disordered Protein States." *Journal of Molecular Biology* 390.3 (2009), pp. 467–477.

22 I. S. Millett, S. Doniach, and K. W. Plaxco. "Toward a Taxonomy of the Denatured State: Small Angle Scattering Studies of Unfolded Proteins." *Advances in Protein Chemistry* 62 (2002), pp. 241–262.

23 P. Bernadó and D. I. Svergun. "Structural Analysis of Intrinsically Disordered Proteins by Small-angle X-ray Scattering." *Molecular Biosystems* 8.1 (2012), pp. 151–167.

24 N. J. Greenfield. "Using Circular Dichroism Spectra to Estimate Protein Secondary Structure." *Nature Protocols* 1.6 (2006), pp. 2876–2890.

25 D. Parker, K. Ferreri, T. Nakajima, V. LaMorte, R. Evans, S. Koerber, C. Hoeger, and M. Montminy. "Phosphorylation of CREB at Ser-133 Induces Complex Formation With CREB-binding Protein via a Direct Mechanism." *Molecular and Cellular Biology* 16.2 (1996), pp. 694–703.

26 M. Brucale, B. Schuler, and B. Samorì. "Single-molecule Studies of Intrinsically Disordered Proteins." *Chemical Reviews* 114.6 (2014), pp. 3281–3317.

27 F. Huang, S. Rajagopalan, G. Settanni, R. J. Marsh, D. A. Armoogum, N. Nicolaou, A. J. Bain, E. Lerner, E. Haas, L. Ying, et al. "Multiple Conformations of Full-length P53 Detected With Single-molecule Fluorescence Resonance Energy Transfer." *Proceedings of the National Academy of Sciences of the United States of America* 106.49 (2009), pp. 20758–20763.

28 J. J. Sakon and K. R. Weninger. "Detecting the Conformation of Individual Proteins in Live Cells." *Nature Methods* 7.3 (2010), pp. 203–205.

29 S. Mukhopadhyay, R. Krishnan, E. A. Lemke, S. Lindquist, and A. A. Deniz. "A Natively Unfolded Yeast Prion Monomer Adopts an Ensemble of Collapsed and Rapidly Fluctuating Structures." *Proceedings of the National Academy of Sciences of the United States of America* 104.8 (2007), pp. 2649–2654.

30 N. Kodera, D. Yamamoto, R. Ishikawa, and T. Ando. "Video Imaging of Walking Myosin v by High-speed Atomic Force Microscopy." *Nature* 468.7320 (2010), pp. 72–76.

31 R. Beveridge, Q. Chappuis, C. Macphee, and P. Barran. "Mass Spectrometry Methods for Intrinsically Disordered Proteins." *Analyst* 138.1 (2013), pp. 32–42.

32 S. Hubbard, R. Beynon, and J. Thornton. "Assessment of Conformational Parameters as Predictors of Limited Proteolytic Sites in Native Protein Structures." *Protein Engineering* 11.5 (1998), pp. 349–359.

33 A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, et al. "Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination." *Acta Crystallographica. Section D, Biological Crystallography* 54.5 (1998), pp. 905–921.

34 A. H. Kwan, M. Mobli, P. R. Gooley, G. F. King, and J. P. Mackay. "Macromolecular NMR Spectroscopy for the Non-spectroscopist." *FEBS Journal* 278.5 (2011), pp. 687–703.

35 C. K. Fisher and C. M. Stultz. "Constructing Ensembles for Intrinsically Disordered Proteins." *Current Opinion in Structural Biology* 21.3 (2011), pp. 426–431.

36 K. Lindorff-Larsen, R. B. Best, M. A. DePristo, C. M. Dobson, and M. Vendruscolo. "Simultaneous Determination of Protein Structure and Dynamics." *Nature* 433.7022 (2005), pp. 128–132.

37 W. F. van Gunsteren, J. Dolenc, and A. E. Mark. "Molecular Simulation as an Aid to Experimentalists." *Current Opinion in Structural Biology* 18.2 (2008), pp. 149–153.

38 S. Rauscher and R. Pomès. "Molecular Simulations of Protein Disorder." *Biochemistry and Cell Biology-Biochimie et Biologie Cellulaire* 88.2 (2010), pp. 269–290.

39 M. Vendruscolo. "Determination of Conformationally Heterogeneous States of Proteins." *Current Opinion in Structural Biology* 17.1 (2007), pp. 15–20.

40 N. G. Sgourakis, Y. Yan, S. A. McCallum, C. Wang, and A. E. Garcia. "The Alzheimer's Peptides Aβ40 and 42 Adopt Distinct Conformations in Water: A Combined MD/NMR Study." *Journal of Molecular Biology* 368.5 (2007), pp. 1448–1457.

41 S. Rauscher, C. Neale, and R. Pomes. "Simulated Tempering Distributed Replica Sampling, Virtual Replica Exchange, and Other Generalized-ensemble Methods for Conformational Sampling." *Journal of Chemical Theory and Computation* 5.10 (2009), pp. 2640–2662.

42 A. Vitalis and R. V. Pappu. "ABSINTH: A New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions." *Journal of Computational Chemistry* 30.5 (2009), pp. 673–699.

43 S. Abeln and D. Frenkel. "Disordered Flanks Prevent Peptide Aggregation." *PLoS Computational Biology* 4.12 (2008), e1000241.

44 J.-M. Mouillon, S. K. Eriksson, and P. Harryson. "Mimicking the Plant Cell Interior Under Water Stress by Macromolecular Crowding: Disordered Dehydrin Proteins Are Highly Resistant to Structural Collapse." *Plant Physiology* 148.4 (2008), pp. 1925–1937.

45 C. Szasz, A. Alexa, K. Toth, M. Rakacs, J. Langowski, and P. Tompa. "Protein Disorder Prevails Under Crowded Conditions." *Biochemistry* 50.26 (2011), pp. 5834–5844.

46 P. Tsvetkov, G. Asher, A. Paz, N. Reuven, J. L. Sussman, I. Silman, and Y. Shaul. "Operational Definition of Intrinsically Unstructured Protein Sequences Based on Susceptibility to the 20S Proteasome." *Proteins: Structure, Function, and Bioinformatics* 70.4 (2008), pp. 1357–1366.

47 J.-F. Bodart, J.-M. Wieruszeski, L. Amniai, A. Leroy, I. Landrieu, A. Rousseau-Lescuyer, J.-P. Vilain, and G. Lippens. "NMR Observation of Tau in Xenopus Oocytes." *Journal of Magnetic Resonance* 192.2 (2008), pp. 252–257.

48 B. C. McNulty, G. B. Young, and G. J. Pielak. "Macromolecular Crowding in the Escherichia Coli Periplasm Maintains α-synuclein Disorder." *Journal of Molecular Biology* 355.5 (2006), pp. 893–897.

49 V. N. Uversky, C. J. Oldfield, and A. K. Dunker. "Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept." *Annual Review of Biophysics* 37 (2008), pp. 215–246.

50 H. Xie, S. Vucetic, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, Z. Obradovic, and V. N. Uversky. "Functional Anthology of Intrinsic Disorder. 3. Ligands, Post-translational Modifications, and Diseases Associated With Intrinsically Disordered Proteins." *Journal of Proteome Research* 6.5 (2007), pp. 1917–1932.

51 L. M. Iakoucheva, C. J. Brown, J. D. Lawson, Z. Obradović, and A. K. Dunker. "Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins." *Journal of Molecular Biology* 323.3 (2002), pp. 573–584.

52 K. H. Vousden and X. Lu. "Live or Let Die: The Cell's Response to p53." *Nature Reviews. Cancer* 2.8 (2002), pp. 594–604.

53 C. J. Oldfield, J. Meng, J. Y. Yang, V. N. Uversky, and A. K. Dunker. "Intrinsic Disorder in Protein-Protein Interaction Networks: Case Studies of Complexes Involving P53 and 14-3-3." In: *The 2007 International Conference on Bioinformatics and Computational Biology*. Vol. 7. CSREA Press, 2007, pp. 553–566.

54 M. Hollstein, D. Sidransky, B. Vogelstein, and C. Harris. "P53 Mutations in Human Cancers." *Science* 253.5015 (1991), pp. 49–53.

55 M. D. Kirkitadze, M. M. Condron, and D. B. Teplow. "Identification and Characterization of Key Kinetic Intermediates in Amyloid β-protein Fibrillogenesis." *Journal of Molecular Biology* 312.5 (2001), pp. 1103–1119.

56 L. K. Simmons, P. C. May, K. J. Tomaselli, R. E. Rydel, K. S. Fuson, E. F. Brigham, S. Wright, I. Lieberburg, G. W. Becker, and D. N. Brems. "Secondary Structure of Amyloid Beta Peptide Correlates With Neurotoxic Activity in Vitro." *Molecular Pharmacology* 45.3 (1994), pp. 373–379.

57 A. K. Dunker and V. N. Uversky. "Signal Transduction via Unstructured Protein Conduits." *Nature Chemical Biology* 4.4 (2008), pp. 229–230.

58 H. J. Dyson and P. E. Wright. "Intrinsically Unstructured Proteins and Their Functions." *Nature Reviews. Molecular Cell Biology* 6.3 (2005), pp. 197–208.

59 J. Liu, N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky, and A. K. Dunker. "Intrinsic Disorder in Transcription Factors." *Biochemistry* 45.22 (2006), pp. 6873–6888.

60 M. Fuxreiter, P. Tompa, I. Simon, V. N. Uversky, J. C. Hansen, and F. J. Asturias. "Malleable Machines Take Shape in Eukaryotic Transcriptional Regulation." *Nature Chemical Biology* 4.12 (2008), pp. 728–737.

61 V. N. Uversky, C. J. Oldfield, and A. K. Dunker. "Showing Your ID: Intrinsic Disorder as an ID for Recognition, Regulation and Cell Signaling." *Journal of Molecular Recognition* 18.5 (2005), pp. 343–384.

62 H. J. Dyson and P. E. Wright. "Coupling of Folding and Binding for Unstructured Proteins." *Current Opinion in Structural Biology* 12.1 (2002), pp. 54–60.

63 H. Xie, S. Vucetic, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, Z. Obradovic, and V. N. Uversky. "Functional Anthology of Intrinsic Disorder. 3. Ligands, Post-translational Modifications, and Diseases Associated With Intrinsically Disordered Proteins." *Journal of Proteome Research* 6.5 (2007), pp. 1917–1932.

64 B. A. Shoemaker, J. J. Portman, and P. G. Wolynes. "Speeding Molecular Recognition by Using the Folding Funnel: The Fly-casting Mechanism." *Proceedings of the National Academy of Sciences of the United States of America* 97.16 (2000), pp. 8868–8873.

65 Y. Huang and Z. Liu. "Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding–binding Process: A Critical Assessment of the "fly-casting" Mechanism." *Journal of Molecular Biology* 393.5 (2009), pp. 1143–1159.

66 C. Haynes, C. J. Oldfield, F. Ji, N. Klitgord, M. E. Cusick, P. Radivojac, V. N. Uversky, M. Vidal, and L. M. Iakoucheva. "Intrinsic Disorder Is a Common Feature of Hub Proteins From Four Eukaryotic Interactomes." *PLoS Computational Biology* 2.8 (2006), e100.

67 R. Van Der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, et al. "Classification of Intrinsically Disordered Regions and Proteins." *Chemical Reviews* 114.13 (2014), pp. 6589–6631.

68 P. Tompa. "The Interplay Between Structure and Function in Intrinsically Unstructured Proteins." *FEBS Letters* 579.15 (2005), pp. 3346–3354.

69 M. O. Collins, L. Yu, I. Campuzano, S. G. Grant, and J. S. Choudhary. "Phosphoproteomic Analysis of the Mouse Brain Cytosol Reveals a Predominance of

Protein Phosphorylation in Regions of Intrinsic Sequence Disorder." *Molecular and Cellular Proteomics* 7.7 (2008), pp. 1331–1348.

70 F. Diella, N. Haslam, C. Chica, A. Budd, S. Michael, N. P. Brown, G. Travé, and T. J. Gibson. "Understanding Eukaryotic Linear Motifs and Their Role in Cell Signaling and Regulation." *Frontiers in Bioscience* 13 (2008), pp. 6580–6603.

71 P. E. Wright and H. J. Dyson. "Linking Folding and Binding." *Current Opinion in Structural Biology* 19.1 (2009), pp. 31–38.

72 B. Xue, P. R. Romero, M. Noutsou, M. M. Maurice, S. G. Rüdiger, A. M. William, M. J. Mizianty, L. Kurgan, V. N. Uversky, and A. K. Dunker. "Stochastic Machines as a Colocalization Mechanism for Scaffold Protein Function." *FEBS Letters* 587.11 (2013), pp. 1587–1591.

73 C. Holt. "Unfolded Phosphopolypeptides Enable Soft and Hard Tissues to Coexist in the Same Organism With Relative Ease." *Current Opinion in Structural Biology* 23.3 (2013), pp. 420–425.

74 N. E. Davey, K. Van Roey, R. J. Weatheritt, G. Toedt, B. Uyar, B. Altenberg, A. Budd, F. Diella, H. Dinkel, and T. J. Gibson. "Attributes of Short Linear Motifs." *Molecular Biosystems* 8.1 (2012), pp. 268–281.

75 K. Van Roey, B. Uyar, R. J. Weatheritt, H. Dinkel, M. Seiler, A. Budd, T. J. Gibson, and N. E. Davey. "Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation." *Chemical Reviews* 114.13 (2014), pp. 6733–6778.

76 C. Pop and G. S. Salvesen. "Human Caspases: Activation, Specificity, and Regulation." *Journal of Biological Chemistry* 284.33 (2009), pp. 21777–21781.

77 J. Pines. "Cyclins and Cyclin-dependent Kinases: A Biochemical View." *Biochemical Journal* 308.Pt 3 (1995), p. 697.

78 T. Pawson and P. Nash. "Assembly of Cell Regulatory Systems Through Protein Interaction Domains." *Science* 300.5618 (2003), pp. 445–452.

79 C. M. Pfleger and M. W. Kirschner. "The KEN Box: An APC Recognition Signal Distinct From the D Box Targeted by Cdh1." *Genes and Development* 14.6 (2000), pp. 655–665.

80 D. Kalderon, B. L. Roberts, W. D. Richardson, and A. E. Smith. "A Short Amino Acid Sequence Able to Specify Nuclear Location." *Cell* 39.3 (1984), pp. 499–509.

81 C. J. Oldfield, Y. Cheng, M. S. Cortese, P. Romero, V. N. Uversky, and A. K. Dunker. "Coupled Folding and Binding With α-helix-forming Molecular Recognition Elements." *Biochemistry* 44.37 (2005), pp. 12454–12470.

82 A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker, and V. N. Uversky. "Analysis of Molecular Recognition Features (MoRFs)." *Journal of Molecular Biology* 362.5 (2006), pp. 1043–1059.

83 M. Fuxreiter, I. Simon, P. Friedrich, and P. Tompa. "Preformed Structural Elements Feature in Partner Recognition by Intrinsically Unstructured Proteins." *Journal of Molecular Biology* 338.5 (2004), pp. 1015–1026.

84 J. W. Chen, P. Romero, V. N. Uversky, and A. K. Dunker. "Conservation of Intrinsic Disorder in Protein Domains and Families: II. Functions of Conserved Disorder." *Journal of Proteome Research* 5.4 (2006), pp. 888–898.

85 M. M. Pentony and D. T. Jones. "Modularity of Intrinsic Disorder in the Human Proteome." *Proteins: Structure, Function, and Bioinformatics* 78.1 (2010), pp. 212–221.

86 D. L. Black. "Mechanisms of Alternative Pre-messenger RNA Splicing." *Annual Review of Biochemistry* 72.1 (2003), pp. 291–336.

87 E. S. Witze, W. M. Old, K. A. Resing, and N. G. Ahn. "Mapping Protein Post-translational Modifications With Mass Spectrometry." *Nature Methods* 4.10 (2007), pp. 798–806.

88 J. Habchi, P. Tompa, S. Longhi, and V. N. Uversky. "Introducing Protein Intrinsic Disorder." *Chemical Reviews* 114.13 (2014), pp. 6561–6588.

89 C. T. Walsh, S. Garneau-Tsodikova, and G. J. Gatto. "Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications." *Angewandte Chemie International Edition* 44.45 (2005), pp. 7342–7372.

90 M. Paetzel, A. Karla, N. C. Strynadka, and R. E. Dalbey. "Signal Peptidases." *Chemical Reviews* 102.12 (2002), pp. 4549–4580.

91 A. Helenius and M. Aebi. "Roles of N-linked Glycans in the Endoplasmic Reticulum." *Annual Review of Biochemistry* 73.1 (2004), pp. 1019–1049.

92 B. M. Turner. "Cellular Memory and the Histone Code." *Cell* 111.3 (2002), pp. 285–291.

93 M. D. Resh. "Fatty Acylation of Proteins: New Insights Into Membrane Targeting of Myristoylated and Palmitoylated Proteins." *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1451.1 (1999), pp. 1–16.

94 C. L. Brooks and W. Gu. "Ubiquitination, Phosphorylation and Acetylation: The Molecular Basis for p53 Regulation." *Current Opinion in Cell Biology* 15.2 (2003), pp. 164–171.

95 G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. "The Protein Kinase Complement of the Human Genome." *Science* 298.5600 (2002), pp. 1912–1934.

96 M. Mann, S.-E. Ong, M. Grønborg, H. Steen, O. N. Jensen, and A. Pandey. "Analysis of Protein Phosphorylation Using Mass Spectrometry: Deciphering the Phosphoproteome." *Trends in Biotechnology* 20.6 (2002), pp. 261–268.

97 J. A. Hoch and T. J. Silhavy. *Two-component Signal Transduction*. Vol. 2. ASM press Washington, DC:, 1995.

98 J. B. Shabb. "Physiological Substrates of cAMP-dependent Protein Kinase." *Chemical Reviews* 101.8 (2001), pp. 2381–2412.

99 A. Alonso, J. Sasin, N. Bottini, I. Friedberg, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon, and T. Mustelin. "Protein Tyrosine Phosphatases in the Human Genome." *Cell* 117.6 (2004), pp. 699–711.

100 J. Hurley, A. Dean, P. Thorsness, D. Koshland, and R. Stroud. "Regulation of Isocitrate Dehydrogenase by Phosphorylation Involves No Long-range Conformational Change in the Free Enzyme." *Journal of Biological Chemistry* 265.7 (1990), pp. 3599–3602.

101 P. Filippakopoulos, M. Kofler, O. Hantschel, G. D. Gish, F. Grebien, E. Salah, P. Neudecker, L. E. Kay, B. E. Turk, G. Superti-Furga, et al. "Structural Coupling

of SH2-kinase Domains Links Fes and Abl Substrate Recognition and Kinase Activation." *Cell* 134.5 (2008), pp. 793–803.

102 C. Peacock and G. Nickless. "The Dissociation Constants of Some Phosphorus (V) Acids." *Zeitschrift für Naturforschung A* 24.2 (1969), pp. 245–247.

103 L. N. Johnson. "The Regulation of Protein Phosphorylation." *Biochemical Society Transactions* 37.4 (2009), pp. 627–641.

104 D. M. Mitrea, C. R. Grace, M. Buljan, M.-K. Yun, N. J. Pytel, J. Satumba, A. Nourse, C.-G. Park, M. M. Babu, S. W. White, et al. "Structural Polymorphism in the N-terminal Oligomerization Domain of NPM1." *Proceedings of the National Academy of Sciences of the United States of America* 111.12 (2014), pp. 4466–4471.

105 M. A. Pufall, G. M. Lee, M. L. Nelson, H.-S. Kang, A. Velyvis, L. E. Kay, L. P. McIntosh, and B. J. Graves. "Variable Control of Ets-1 DNA Binding by Multiple Phosphates in an Unstructured Region." *Science* 309.5731 (2005), pp. 142–145.

106 R. Bomblies, M. Luitz, and M. Zacharias. "Characterization of Transiently Stable Structural Motifs in Intrinsically Disordered Proteins Using Free Energy Simulations." *Biophysical Journal* 108.2 (2015), 317a–318a.

107 L. M. Espinoza-Fonseca, D. Kast, and D. D. Thomas. "Thermodynamic and Structural Basis of Phosphorylation-induced Disorder-to-order Transition in the Regulatory Light Chain of Smooth Muscle Myosin." *Journal of the American Chemical Society* 130.37 (2008), pp. 12208–12209.

108 J. L. Smart and J. A. McCammon. "Phosphorylation Stabilizes the N-termini of α-helices." *Biopolymers* 49.3 (1999), pp. 225–233.

109 D. Frenkel and B. Smit. *Understanding Molecular Simulation*. 2nd. Orlando, FL, USA: Academic Press, Inc., 2001.

110 J. E. Jones. "On the Determination of Molecular Fields. II. From the Equation of State of a Gas." In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. Vol. 106. 738. The Royal Society. 1924, pp. 463–477.

111 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules." *Journal of the American Chemical Society* 117.19 (1995), pp. 5179–5197.

112 D. Case, T. Darden, T. Cheatham III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang, K. Merz, et al. "AMBER 12." *University of California, San Francisco* (2012).

113 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, et al. "CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible With the CHARMM All-atom Additive Biological Force Fields." *Journal of Computational Chemistry* 31.4 (2010), pp. 671–690.

114 M. M. Reif, M. Winger, and C. Oostenbrink. "Testing of the GROMOS Forcefield Parameter Set 54A8: Structural Properties of Electrolyte Solutions, Lipid Bilayers, and Proteins." *Journal of Chemical Theory and Computation* 9.2 (2013), pp. 1247–1264.

115 L. Verlet. “Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules.” *Physical Review* 159.1 (1967), p. 98.

116 W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. “A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters.” *Journal of Chemical Physics* 76.1 (1982), pp. 637–649.

117 R. W. Hockney. *Potential Calculation and Some Applications*. Tech. rep. Langley Research Center, Hampton, Va., 1970.

118 J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen. “Numerical Integration of the Cartesian Equations of Motion of a System With Constraints: Molecular Dynamics of N-alkanes.” *Journal of Computational Physics* 23.3 (1977), pp. 327–341.

119 C. W. Hopkins, S. Le Grand, R. C. Walker, and A. E. Roitberg. “Long-time-step Molecular Dynamics Through Hydrogen Mass Repartitioning.” *Journal of Chemical Theory and Computation* 11.4 (2015), pp. 1864–1874.

120 H. J. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. Haak. “Molecular Dynamics With Coupling to an External Bath.” *Journal of Chemical Physics* 81.8 (1984), pp. 3684–3690.

121 W. G. Hoover. “Canonical Dynamics: Equilibrium Phase-space Distributions.” *Physical Review A* 31.3 (1985), p. 1695.

122 G. J. Martyna, M. L. Klein, and M. Tuckerman. “Nosé–Hoover chains: the canonical ensemble via continuous dynamics.” *Journal of Chemical Physics* 97.4 (1992), pp. 2635–2643.

123 P. H. Hünenberger. “Thermostat Algorithms for Molecular Dynamics Simulations.” In: *Advanced Computer Simulation*. Springer, 2005, pp. 105–149.

124 H. C. Andersen. “Molecular Dynamics Simulations at Constant Pressure and/or Temperature.” *Journal of Chemical Physics* 72.4 (1980), pp. 2384–2393.

125 M. Parrinello and A. Rahman. “Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method.” *Journal of Applied Physics* 52.12 (1981), pp. 7182–7190.

126 N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. “Equation of State Calculations by Fast Computing Machines.” *Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092.

127 J. A. McCammon, B. R. Gelin, and M. Karplus. “Dynamics of Folded Proteins.” *Nature* 267.5612 (1977), pp. 585–590.

128 K. Toukan and A. Rahman. “Molecular-dynamics Study of Atomic Motions in Water.” *Physical Review B* 31.5 (1985), p. 2643.

129 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. “Comparison of Simple Potential Functions for Simulating Liquid Water.” *Journal of Chemical Physics* 79.2 (1983), pp. 926–935.

130 J. L. Abascal and C. Vega. “A general purpose model for the condensed phases of water: TIP4P/2005.” *The Journal of chemical physics* 123.23 (2005), p. 234505.

131 W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson. “Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics.” *Journal of the American Chemical Society* 112.16 (1990), pp. 6127–6129.

132 B. Lu, D. Zhang, and J. A. McCammon. "Computation of Electrostatic Forces Between Solvated Molecules Determined by the Poisson–Boltzmann Equation Using a Boundary Element Method." *Journal of Chemical Physics* 122.21 (2005), p. 214102.

133 P. Koehl. "Electrostatics Calculations: Latest Methodological Advances." *Current Opinion in Structural Biology* 16.2 (2006), pp. 142–151.

134 I. Grimlock. commons.wikimedia.org/w/index.php?curid=2520550, CC BY-SA 3.0. 2007.

135 I. Ohmine, H. Tanaka, and P. G. Wolynes. "Large Local Energy Fluctuations in Water. II. Cooperative Motions and Fluctuations." *Journal of Chemical Physics* 89.9 (1988), pp. 5852–5860.

136 T. Darden, D. York, and L. Pedersen. "Particle Mesh Ewald: An N Log (N) Method for Ewald Sums in Large Systems." *Journal of Chemical Physics* 98.12 (1993), pp. 10089–10092.

137 P. P. Ewald. "Die Berechnung Optischer Und Elektrostatischer Gitterpotentiale." *Annals of Physics* 369.3 (1921), pp. 253–287.

138 D. Li and R. Brüschweiler. "PPM_One: A Static Protein Structure Based Chemical Shift Predictor." *Journal of Biomolecular NMR* 62.3 (2015), pp. 403–409.

139 D. Svergun, C. Barberato, and M. Koch. "CRYSOL–a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules From Atomic Coordinates." *Journal of Applied Crystallography* 28.6 (1995), pp. 768–773.

140 M. Hoefling and H. Grubmüller. "In Silico FRET From Simulated Dye Dynamics." *Computer Physics Communications* 184.3 (2013), pp. 841–852.

141 C. N. Cavasotto, R. Bomblies, M. Luitz, and M. Zacharias. "Free Energy Calculations of Ligand–Protein Binding." In: *In Silico Drug Discovery and Design: Theory, Methods, Challenges, and Applications*. CRC Press, 2015, pp. 313–335.

142 H. Fujitani, Y. Tanida, M. Ito, G. Jayachandran, C. D. Snow, M. R. Shirts, E. J. Sorin, and V. S. Pande. "Direct Calculation of the Binding Free Energies of FKBP Ligands." *Journal of Chemical Physics* 123.8 (2005).

143 M. R. Shirts. "Best Practices in Free Energy Calculations for Drug Design." In: *Computational Drug Discovery and Design*. Springer New York, 2012, pp. 425–467.

144 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. "Development and Testing of a General Amber Force Field." *Journal of Computational Chemistry* 25.9 (2004), pp. 1157–1174.

145 R. W. Zwanzig. "High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases." *Journal of Chemical Physics* 22.8 (1954), pp. 1420–1426.

146 C. H. Bennett. "Efficient Estimation of Free Energy Differences From Monte Carlo Data." *Journal of Computational Physics* 22 (1976), pp. 245–268.

147 A. de Ruiter, S. Boresch, and C. Oostenbrink. "Comparison of Thermodynamic Integration and Bennett Acceptance Ratio for Calculating Relative Protein-ligand Binding Free Energies." *Journal of Computational Chemistry* 34.12 (2013), pp. 1024–1034.

148 J. A. Barker and D. Henderson. "What Is "liquid"? Understanding the States of Matter." *Reviews of Modern Physics* 48 (4 Oct. 1976), pp. 587–671.

149 D. L. Beveridge and F. M. DiCapua. "Free Energy via Molecular Simulation: Applications to Chemical and Biomolecular Systems." *Annual Review of Biophysics and Biophysical Chemistry* 18.1 (1989), pp. 431–492.

150 J. G. Kirkwood. "Statistical Mechanics of Fluid Mixtures." *Journal of Chemical Physics* 3.5 (1935), pp. 300–313.

151 S. Bruckner and S. Boresch. "Efficiency of Alchemical Free Energy Simulations. II. Improvements for Thermodynamic Integration." *Journal of Computational Chemistry* 32.7 (2011), pp. 1320–1333.

152 M. Zacharias, T. P. Straatsma, and J. A. McCammon. "Separation–shifted Scaling, a New Scaling Method for Lennard–Jones Interactions in Thermodynamic Integration." *Journal of Chemical Physics* 100.12 (1994), pp. 9025–9031.

153 T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren. "Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations." *Chemical Physics Letters* 222 (1994), pp. 529–539.

154 V. Gapsys, D. Seeliger, and B. L. de Groot. "New Soft-Core Potential Function for Molecular Dynamics Based Alchemical Free Energy Calculations." *Journal of Chemical Theory and Computation* 8.7 (2012), pp. 2373–2382.

155 T. T. Pham and M. R. Shirts. "Identifying Low Variance Pathways for Free Energy Calculations of Molecular Transformations in Solution Phase." *Journal of Chemical Physics* 135.3 (2011).

156 D. Chandler. *Introduction to modern statistical mechanics*. New York, Oxford: Oxford University Press, 1987.

157 J. Hénin and C. Chipot. "Overcoming Free Energy Barriers Using Unconstrained Molecular Dynamics Simulations." *Journal of Chemical Physics* 121.7 (2004), pp. 2904–2914.

158 E. Darve and A. Pohorille. "Calculating Free Energies Using Average Force." *Journal of Chemical Physics* 115.20 (2001), pp. 9169–9183.

159 W. K. den Otter and W. J. Briels. "The Calculation of Free-energy Differences by Constrained Molecular-dynamics Simulations." *Journal of Chemical Physics* 109.11 (1998), pp. 4139–4146.

160 B. Roux. "The Calculation of the Potential of Mean Force Using Computer Simulations." *Computer Physics Communications* 91.1–3 (1995), pp. 275–282.

161 S. Doudou, N. A. Burton, and R. H. Henchman. "Standard Free Energy of Binding From a One-Dimensional Potential of Mean Force." *Journal of Chemical Theory and Computation* 5.4 (2009), pp. 909–918.

162 G. M. Torrie and J. P. Valleau. "Monte Carlo Free Energy Estimates Using Non-Boltzmann Sampling: Application to the Sub-critical Lennard-Jones Fluid." *Chemical Physics Letters* 28.4 (1974), pp. 578–581.

163 G. N. Patey and J. P. Valleau. "A Monte Carlo Method for Obtaining the Interionic Potential of Mean Force in Ionic Solution." *Journal of Chemical Physics* 63.6 (1975), pp. 2334–2339.

164 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. "The Weighted Histogram Analysis Method for Free-energy Calculations on

Biomolecules. I. The Method." *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021.

165 J. Kästner and W. Thiel. "Analysis of the Statistical Error in Umbrella Sampling Simulations by Umbrella Integration." *Journal of Chemical Physics* 124.23 (2006).

166 T.-S. Lee, B. K. Radak, A. Pabis, and D. M. York. "A New Maximum Likelihood Approach for Free Energy Profile Construction From Molecular Simulations." *Journal of Chemical Theory and Computation* 9.1 (2013), pp. 153–164.

167 H.-J. Woo and B. Roux. "Calculation of Absolute Protein–ligand Binding Free Energy From Computer Simulations." *Proc. Natl. Acad. Sci. U.S.A.* 102.19 (2005), pp. 6825–6830.

168 Y. Deng and B. Roux. "Computations of Standard Binding Free Energies With Molecular Dynamics Simulations." *Journal of Physical Chemistry B* 113.8 (2009), pp. 2234–2246.

169 W. Jiang and B. Roux. "Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations." *Journal of Chemical Theory and Computation* 6.9 (2010), pp. 2559–2565.

170 W. Jiang, Y. Luo, L. Maragliano, and B. Roux. "Calculation of Free Energy Landscape in Multi-Dimensions With Hamiltonian-Exchange Umbrella Sampling on Petascale Supercomputer." *Journal of Chemical Theory and Computation* 8.11 (2012), pp. 4672–4680.

171 J. C. Gumbart, B. Roux, and C. Chipot. "Standard Binding Free Energies From Computer Simulations: What Is the Best Strategy?" *Journal of Chemical Theory and Computation* 9.1 (2012), pp. 794–802.

172 K. Huang and A. E. Garcia. "Free Energy of Translocating an Arginine-Rich Cell-Penetrating Peptide Across a Lipid Bilayer Suggests Pore Formation." *Biophysical Journal* 104.2 (2013), pp. 412–420.

173 C. Velez-Vega and M. K. Gilson. "Overcoming Dissipation in the Calculation of Standard Binding Free Energies by Ligand Extraction." *Journal of Computational Chemistry* 34.27 (2013), pp. 2360–2371.

174 F. Zeller and M. Zacharias. "Adaptive Biasing Combined With Hamiltonian Replica Exchange to Improve Umbrella Sampling Free Energy Simulations." *Journal of Chemical Theory and Computation* 10 (2014), pp. 703–710.

175 H. Liu, A. E. Mark, and W. F. van Gunsteren. "Estimating the Relative Free Energy of Different Molecular States With Respect to a Single Reference State." *Journal of Physical Chemistry* 100.22 (1996), pp. 9485–9494.

176 C. Oostenbrink and W. F. van Gunsteren. "Free Energies of Ligand Binding for Structurally Diverse Compounds." *Proc. Natl. Acad. Sci. U.S.A.* 102.19 (2005), pp. 6750–6754.

177 J. Hritz, T. Läppchen, and C. Oostenbrink. "Calculations of Binding Affinity Between C8-substituted GTP Analogs and the Bacterial Cell-division Protein FtsZ." *European Biophysics Journal* 39.12 (2010), pp. 1573–1580.

178 C. Oostenbrink. "Free Energy Calculations From One-Step Perturbations." In: *Computational Drug Discovery and Design*. Springer New York, 2012, pp. 487–499.

179 C. Jarzynski. "Nonequilibrium Equality for Free Energy Differences." *Physical Review Letters* 78 (14 Apr. 1997), pp. 2690–2693.

180 C. Oostenbrink and W. F. van Gunsteren. "Calculating Zeros: Non-equilibrium Free Energy Calculations." *Chemical Physics* 323.1 (2006), pp. 102–108.

181 H. Oberhofer, C. Dellago, and P. L. Geissler. "Biased Sampling of Nonequilibrium Trajectories: Can Fast Switching Simulations Outperform Conventional Free Energy Calculation Methods." *Journal of Physical Chemistry B* 109.14 (2005), pp. 6902–6915.

182 D. Wu and D. A. Kofke. "Phase-space Overlap Measures. II. Design and Implementation of Staging Methods for Free-energy Calculations." *Journal of Chemical Physics* 123 (2005), p. 84109.

183 B. R. Brooks, C. L. Brooks, A. D. MacKerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al. "CHARMM: The Biomolecular Simulation Program." *Journal of Computational Chemistry* 30.10 (2009), pp. 1545–1614.

184 S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl. "GROMACS 4.5: A High-throughput and Highly Parallel Open Source Molecular Simulation Toolkit." *Bioinformatics* 29.7 (2013), pp. 845–854.

185 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. "Scalable Molecular Dynamics With NAMD." *Journal of Computational Chemistry* 26.16 (2005), pp. 1781–1802.

186 J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson, and V. S. Pande. "Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges." *Current Opinion in Structural Biology* 21.2 (2011), pp. 150–160.

187 S. Jo, W. Jiang, H. S. Lee, B. Roux, and W. Im. "CHARMM-GUI Ligand Binder for Absolute Binding Free Energy Calculations and Its Application." *Journal of Chemical Information and Modeling* 53.1 (2013), pp. 267–277.

188 C. D. Christ and T. Fox. "Accuracy Assessment and Automation of Free Energy Calculations for Drug Design." *Journal of Chemical Information and Modeling* 54.1 (2014), pp. 108–120.

189 **R. Bomblies**, M. P. Luitz, and M. Zacharias. "Mechanism of pKID/KIX Association Studied by Molecular Dynamics Free Energy Simulations." *Journal of Physical Chemistry B* 120.33 (Apr. 2016), pp. 8186–8192.

190 K. Ostermeir and M. Zacharias. "Advanced Replica-exchange Sampling to Study the Flexibility and Plasticity of Peptides and Proteins." *Biochimica et Biophysica Acta, Proteins and Proteomics* 1834.5 (2013), pp. 847–853.

191 C. J. Woods, J. W. Essex, and M. A. King. "The Development of Replica-Exchange-Based Free-Energy Methods." *Journal of Physical Chemistry B* 107.49 (2003), pp. 13703–13710.

192 M. P. Luitz and M. Zacharias. "Role of Tyrosine Hot-spot Residues at the Interface of Colicin E9 and Immunity Protein 9: A Comparative Free Energy Simulation Study." *Proteins: Structure, Function, and Bioinformatics* 81.3 (2013), pp. 461–468.

193 H. Kokubo, T. Tanaka, and Y. Okamoto. "Ab Initio Prediction of Protein–Ligand Binding Structures by Replica-exchange Umbrella Sampling Simulations." *Journal of Computational Chemistry* 32.13 (2011), pp. 2810–2821.

194 S. Kannan and M. Zacharias. "Folding Simulations of Trp-cage Mini Protein in Explicit Solvent Using Biasing Potential Replica-exchange Molecular Dynamics Simulations." *Proteins: Structure, Function, and Bioinformatics* 76.2 (2009), pp. 448–460.

195 L. Wang, B. Berne, and R. A. Friesner. "On Achieving High Accuracy and Reliability in the Calculation of Relative Protein–ligand Binding Affinities." *Proceedings of the National Academy of Sciences of the United States of America* 109.6 (2012), pp. 1937–1942.

196 J. L. Knight and C. L. Brooks. "λ-Dynamics Free Energy Simulation Methods." *Journal of Computational Chemistry* 30.11 (2009), pp. 1692–1700.

197 Z. Guo, J. Durkin, T. Fischmann, R. Ingram, A. Prongay, R. Zhang, and V. Madison. "Application of the λ-Dynamics Method to Evaluate the Relative Binding Free Energies of Inhibitors to HCV Protease." *Journal of Medicinal Chemistry* 46.25 (2003), pp. 5360–5364.

198 J. Mongan and D. A. Case. "Biomolecular Simulations at Constant pH." *Current Opinion in Structural Biology* 15.2 (2005), pp. 157–163.

199 A. Laio and M. Parrinello. "Escaping Free-energy Minima." *Proceedings of the National Academy of Sciences of the United States of America* 99.20 (2002), pp. 12562–12566.

200 A. Barducci, G. Bussi, and M. Parrinello. "Well-tempered Metadynamics: A Smoothly Converging and Tunable Free-energy Method." *Physical Review Letters* 100.2 (2008), p. 020603.

201 A. Barducci, M. Bonomi, and M. Parrinello. "Metadynamics." *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.5 (2011), pp. 826–843.

202 V. Limongelli, L. Marinelli, S. Cosconati, C. La Motta, S. Sartini, L. Mugnaini, F. Da Settimo, E. Novellino, and M. Parrinello. "Sampling Protein Motion and Solvent Effect During Ligand Binding." *Proceedings of the National Academy of Sciences of the United States of America* 109.5 (2012), pp. 1467–1472.

203 A. Barducci, R. Chelli, P. Procacci, V. Schettino, F. L. Gervasio, and M. Parrinello. "Metadynamics Simulation of Prion Protein: β-structure Stability and the Early Stages of Misfolding." *Journal of the American Chemical Society* 128.8 (2006), pp. 2705–2710.

204 G. Fiorin, A. Pastore, P. Carloni, and M. Parrinello. "Using Metadynamics to Understand the Mechanism of Calmodulin/target Recognition at Atomic Detail." *Biophysical Journal* 91.8 (2006), pp. 2768–2777.

205 F. L. Gervasio, M. Parrinello, M. Ceccarelli, and M. L. Klein. "Exploring the Gating Mechanism in the ClC Chloride Channel via Metadynamics." *Journal of Molecular Biology* 361.2 (2006), pp. 390–398.

206 V. Babin, C. Roland, and C. Sagui. "Adaptively Biased Molecular Dynamics for Free Energy Calculations." *Journal of Chemical Physics* 128.13 (2008), p. 134101.

207 J. C. Gumbart, B. Roux, and C. Chipot. "Efficient Determination of Protein–protein Standard Binding Free Energies From First Principles." *Journal of Chemical Theory and Computation* 9.8 (2013), pp. 3789–3798.

208 S. Rauscher, V. Gapsys, M. J. Gajda, M. Zweckstetter, B. L. de Groot, and H. Grubmüller. "Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field : A Comparison to Experiment." *Journal of Chemical Theory and Computation* 11.11 (2015), pp. 5513–5524.

209 K. Lindorff-Larsen, N. Trbovic, P. Maragakis, S. Piana, and D. E. Shaw. "Structure and Dynamics of an Unfolded Protein Examined by Molecular Dynamics Simulation." *Journal of the American Chemical Society* 134.8 (2012), pp. 3787–3791.

210 K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw. "How Fast-folding Proteins Fold." *Science* 334.6055 (2011), pp. 517–520.

211 S. Piana, K. Lindorff-Larsen, and D. E. Shaw. "How Robust Are Protein Folding Simulations With Respect to Force Field Parameterization?" *Biophysical Journal* 100.9 (2011), pp. L47–L49.

212 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. "Improved Side-chain Torsion Potentials for the Amber Ff99SB Protein Force Field." *Proteins: Structure, Function, and Bioinformatics* 78.8 (2010), pp. 1950–1958.

213 K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw. "Systematic Validation of Protein Force Fields Against Experimental Data." *PLOS ONE* 7.2 (2012), e32131.

214 K. A. Beauchamp, Y.-S. Lin, R. Das, and V. S. Pande. "Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements." *Journal of Chemical Theory and Computation* 8.4 (2012), pp. 1409–1414.

215 E. A. Cino, W.-Y. Choy, and M. Karttunen. "Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations." *Journal of Chemical Theory and Computation* 8.8 (2012), pp. 2725–2740.

216 E. Yedvabny, P. S. Nerenberg, C. So, and T. Head-Gordon. "Disordered Structural Ensembles of Vasopressin and Oxytocin and Their Mutants." *Journal of Physical Chemistry B* 119.3 (2014), pp. 896–905.

217 N. G. Sgourakis, M. Merced-Serrano, C. Boutsidis, P. Drineas, Z. Du, C. Wang, and A. E. García. "Atomic-level Characterization of the Ensemble of the Aβ (1–42) Monomer in Water Using Unbiased Molecular Dynamics Simulations and Spectral Algorithms." *Journal of Molecular Biology* 405.2 (2011), pp. 570–583.

218 G. A. Pantelopulos, S. Mukherjee, and V. A. Voelz. "Microsecond Simulations of Mdm2 and Its Complex With p53 Yield Insight Into Force Field Accuracy and Conformational Dynamics." *Proteins: Structure, Function, and Bioinformatics* 83.9 (2015), pp. 1665–1676.

219 D. J. Rosenman, C. R. Connors, W. Chen, C. Wang, and A. E. García. "Aβ Monomers Transiently Sample Oligomer and Fibril-like Configurations: Ensemble Characterization Using a Combined MD/NMR Approach." *Journal of Molecular Biology* 425.18 (2013), pp. 3338–3359.

220 A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine, and R. V. Pappu. "Net Charge Per Residue Modulates Conformational Ensembles of Intrinsically Disordered Proteins." *Proceedings of the National Academy of Sciences of the United States of America* 107.18 (2010), pp. 8183–8188.

221 A. M. Fluitt and J. J. de Pablo. "An Analysis of Biomolecular Force Fields for Simulations of Polyglutamine in Solution." *Biophysical Journal* 109.5 (2015), pp. 1009–1018.

222 S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw. "Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States." *Journal of Physical Chemistry B* 119.16 (2015), pp. 5113–5123.

223 W. Ye, D. Ji, W. Wang, R. Luo, and H.-F. Chen. "Test and Evaluation of Ff99IDPs Force Field for Intrinsically Disordered Proteins." *Journal of Chemical Information and Modeling* 55.5 (2015), pp. 1021–1029.

224 K. Q. Hoffmann, M. McGovern, C.-c. Chiu, and J. J. de Pablo. "Secondary Structure of Rat and Human Amylin Across Force Fields." *PLOS ONE* 10.7 (2015), e0134091.

225 R. B. Best and J. Mittal. "Free-energy Landscape of the GB1 Hairpin in All-atom Explicit Solvent Simulations With Different Force Fields: Similarities and Differences." *Proteins: Structure, Function, and Bioinformatics* 79.4 (2011), pp. 1318–1328.

226 S. Piana, J. L. Klepeis, and D. E. Shaw. "Assessing the Accuracy of Physical Models Used in Protein-folding Simulations: Quantitative Evidence From Long Molecular Dynamics Simulations." *Current Opinion in Structural Biology* 24 (2014), pp. 98–105.

227 J. J. Skinner, W. Yu, E. K. Gichana, M. C. Baxa, J. R. Hinshaw, K. F. Freed, and T. R. Sosnick. "Benchmarking All-atom Simulations Using Hydrogen Exchange." *Proceedings of the National Academy of Sciences of the United States of America* 111.45 (2014), pp. 15975–15980.

228 R. B. Best, W. Zheng, and J. Mittal. "Balanced Protein–water Interactions Improve Properties of Disordered Proteins and Non-specific Protein Association." *Journal of Chemical Theory and Computation* 10.11 (2014), pp. 5113–5124.

229 D. Mercadante, S. Milles, G. Fuertes, D. I. Svergun, E. A. Lemke, and F. Gräter. "Kirkwood–Buff Approach Rescues Overcollapse of a Disordered Protein in Canonical Protein Force Fields." *Journal of Physical Chemistry B* 119.25 (2015), pp. 7975–7984.

230 M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande. "Extremely Precise Free Energy Calculations of Amino Acid Side Chain Analogs: Comparison of Common Molecular Mechanics Force Fields for Proteins." *Journal of Chemical Physics* 119.11 (2003), pp. 5740–5761.

231 D. Petrov and B. Zagrovic. "Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments?" *PLoS Computational Biology* 10.5 (2014), e1003638.

232 D. Nettels, S. Müller-Späth, F. Küster, H. Hofmann, D. Haenni, S. Rüegger, L. Reymond, A. Hoffmann, J. Kubelka, B. Heinz, et al. "Single-molecule Spectroscopy of the Temperature-induced Collapse of Unfolded Proteins." *Proceedings of the National Academy of Sciences of the United States of America* 106.49 (2009), pp. 20740–20745.

233 H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon. "Development of an Improved Four-site Water Model for Biomolecular Simulations: TIP4P-Ew." *Journal of Chemical Physics* 120.20 (2004), pp. 9665–9678.

234 T. A. Halgren and W. Damm. "Polarizable Force Fields." *Current Opinion in Structural Biology* 11.2 (2001), pp. 236–242.

235 G. A. Kaminski, H. A. Stern, B. J. Berne, R. A. Friesner, Y. X. Cao, R. B. Murphy, R. Zhou, and T. A. Halgren. "Development of a Polarizable Force Field for Proteins via Ab Initio Quantum Chemistry: First Generation Model and Gas Phase Tests." *Journal of Computational Chemistry* 23.16 (2002), pp. 1515–1531.

236 P. Ren and J. W. Ponder. "Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanics Calculations." *Journal of Computational Chemistry* 23.16 (2002), pp. 1497–1506.

237 J. L. Banks, G. A. Kaminski, R. Zhou, D. T. Mainz, B. J. Berne, and R. A. Friesner. "Parametrizing a Polarizable Force Field From Ab Initio Data. I. The Fluctuating Point Charge Model." *Journal of Chemical Physics* 110.2 (1999), pp. 741–754.

238 M. Karplus and J. A. McCammon. "Molecular Dynamics Simulations of Biomolecules." *Nature Structural and Molecular Biology* 9.9 (2002), pp. 646–652.

239 L. Zeng, F. Fagotto, T. Zhang, W. Hsu, T. J. Vasicek, W. L. Perry, J. J. Lee, S. M. Tilghman, B. M. Gumbiner, and F. Costantini. "The Mouse Fusedlocus Encodes Axin, an Inhibitor of the Wnt Signaling Pathway That Regulates Embryonic Axis Formation." *Cell* 90.1 (1997), pp. 181–192.

240 M. van Noort, J. Meeldijk, R. van der Zee, O. Destree, and H. Clevers. "Wnt Signaling Controls the Phosphorylation Status of β-catenin." *Journal of Biological Chemistry* 277.20 (2002), pp. 17901–17905.

241 L.-K. Su, B. Vogelstein, and K. W. Kinzler. "Association of the APC Tumor Suppressor Protein With Catenins." *Science* 262.5140 (1993), pp. 1734–1737.

242 B. Rubinfeld, B. Souza, I. Albert, O. Muller, S. H. Chamberlain, F. R. Masiarz, S. Munemitsu, and P. Polakis. "Association of the APC Gene Product With Beta-catenin." *Science* 262.5140 (1993), pp. 1731–1734.

243 K. E. Spink, P. Polakis, and W. I. Weis. "Structural Basis of the Axin–adenomatous Polyposis Coli Interaction." *EMBO Journal* 19.10 (2000), pp. 2270–2279.

244 M. Noutsou, A. M. Duarte, Z. Anvarian, T. Didenko, D. P. Minde, I. Kuper, I. De Ridder, C. Oikonomou, A. Friedler, R. Boelens, et al. "Critical Scaffolding Regions of the Tumor Suppressor Axin1 Are Natively Unfolded." *Journal of Molecular Biology* 405.3 (2011), pp. 773–786.

245 Y. Xing, W. K. Clements, D. Kimelman, and W. Xu. "Crystal Structure of a β-catenin/axin Complex Suggests a Mechanism for the β-catenin Destruction Complex." *Genes and Development* 17.22 (2003), pp. 2753–2764.

246 J. L. Stamos and W. I. Weis. "The β-catenin Destruction Complex." *Cold Spring Harbor Perspectives in Biology* 5.1 (2013), a007898.

247 R. Dajani, E. Fraser, S. M. Roe, M. Yeo, V. M. Good, V. Thompson, T. C. Dale, and L. H. Pearl. "Structural Basis for Recruitment of Glycogen Synthase Kinase 3β to the Axin—APC Scaffold Complex." *EMBO Journal* 22.3 (2003), pp. 494–501.

248 B. Xue, A. K. Dunker, and V. N. Uversky. "The Roles of Intrinsic Disorder in Orchestrating the Wnt-pathway." *Journal of Biomolecular Structure and Dynamics* 29.5 (2012), pp. 843–861.

249 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. "Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters." *Proteins: Structure, Function, and Bioinformatics* 65.3 (2006), pp. 712–725.

250 R. B. Best and G. Hummer. "Optimized Molecular Dynamics Force Fields Applied to the Helix- Coil Transition of Polypeptides." *Journal of Physical Chemistry B* 113.26 (2009), pp. 9004–9015.

251 B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation." *Journal of Chemical Theory and Computation* 4.3 (2008), pp. 435–447.

252 T. Darden, D. York, and L. Pedersen. "Particle Mesh Ewald: An N Log (N) Method for Ewald Sums in Large Systems." *Journal of Chemical Physics* 98.12 (1993), pp. 10089–10092.

253 B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije. "LINCS: A Linear Constraint Solver for Molecular Simulations." *Journal of Computational Chemistry* 18.12 (1997), pp. 1463–1472.

254 G. Bussi, D. Donadio, and M. Parrinello. "Canonical Sampling Through Velocity Rescaling." *Journal of Chemical Physics* 126.1 (2007), p. 014101.

255 M. Parrinello and A. Rahman. "Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method." *Journal of Applied Physics* 52.12 (1981), pp. 7182–7190.

256 Schrödinger, LLC. "The PyMOL Molecular Graphics System, Version 1.8." Nov. 2015.

257 F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. "NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes." *Journal of Biomolecular NMR* 6.3 (1995), pp. 277–293.

258 R. Fogh, J. Ionides, E. Ulrich, W. Boucher, W. Vranken, J. P. Linge, M. Habeck, W. Rieping, T. Bhat, J. Westbrook, et al. "The CCPN Project: An Interim Report on a Data Model for the NMR Community." *Nature Structural and Molecular Biology* 9.6 (2002), pp. 416–418.

259 K. Tamiola and F. A. Mulder. "Using NMR Chemical Shifts to Calculate the Propensity for Structural Order and Disorder in Proteins." *Biochemical Society Transactions* 40.5 (2012), pp. 1014–1020.

260 W. Kabsch and C. Sander. "Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-bonded and Geometrical Features." *Biopolymers* 22.12 (1983), pp. 2577–2637.

261 T. Güttler, T. Madl, P. Neumann, D. Deichsel, L. Corsini, T. Monecke, R. Ficner, M. Sattler, and D. Görlich. "NES Consensus Redefined by Structures of PKI-type and Rev-type Nuclear Export Signals Bound to CRM1." *Nature Structural and Molecular Biology* 17.11 (2010), pp. 1367–1376.

262 W. Gu, X.-L. Shi, and R. G. Roeder. "Synergistic Activation of Transcription by CBP and p53." *Nature* 387.6635 (1997), pp. 819–822.

263 N. Nasrin, S. Ogg, C. M. Cahill, W. Biggs, S. Nui, J. Dore, D. Calvo, Y. Shi, G. Ruvkun, and M. C. Alexander-Bridges. "DAF-16 Recruits the CREB-binding Protein Coactivator Complex to the Insulin-like Growth Factor Binding Protein 1

Promoter in HepG2 Cells." *Proceedings of the National Academy of Sciences of the United States of America* 97.19 (2000), pp. 10412–10417.

264 G. A. Blobel, T. Nakajima, R. Eckner, M. Montminy, and S. H. Orkin. "CREB-binding Protein Cooperates With Transcription Factor GATA-1 and Is Required for Erythroid Differentiation." *Proceedings of the National Academy of Sciences of the United States of America* 95.5 (1998), pp. 2061–2066.

265 V. Doucas, M. Tini, D. A. Egan, and R. M. Evans. "Modulation of CREB Binding Protein Function by the Promyelocytic (PML) Oncoprotein Suggests a Role for Nuclear Bodies in Hormone Signaling." *Proceedings of the National Academy of Sciences of the United States of America* 96.6 (1999), pp. 2627–2632.

266 I. Radhakrishnan, G. C. Pérez-Alvarado, H. J. Dyson, and P. E. Wright. "Conformational Preferences in the Ser133 - Phosphorylated and Non - Phosphorylated Forms of the Kinase Inducible Transactivation Domain of CREB." *FEBS Letters* 430.3 (1998), pp. 317–322.

267 D. Parker, M. Rivera, T. Zor, A. Henrion-Caude, I. Radhakrishnan, A. Kumar, L. H. Shapiro, P. E. Wright, M. Montminy, and P. K. Brindle. "Role of Secondary Structure in Discrimination Between Constitutive and Inducible Activators." *Molecular and Cellular Biology* 19.8 (1999), pp. 5601–5607.

268 T. Zor, B. M. Mayr, H. J. Dyson, M. R. Montminy, and P. E. Wright. "Roles of Phosphorylation and Helix Propensity in the Binding of the KIX Domain of CREB-binding Protein by Constitutive (c-Myb) and Inducible (CREB) Activators." *Journal of Biological Chemistry* 277.44 (2002), pp. 42241–42248.

269 S. P. Mestas and K. J. Lumb. "Electrostatic Contribution of Phosphorylation to the Stability of the CREB–CBP Activator-Coactivator Complex." *Nature Structural and Molecular Biology* 6.7 (1999), pp. 613–614.

270 I. Radhakrishnan, G. C. Pérez-Alvarado, D. Parker, H. J. Dyson, M. R. Montminy, and P. E. Wright. "Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of CREB: A Model for Activator:Coactivator Interactions." *Cell* 91.6 (1997), pp. 741–752.

271 M. Dal Peraro, F. Alber, and P. Carloni. "Ser133 Phosphate-KIX Interactions in the CREB-CBP Complex: An Ab Initio Molecular Dynamics Study." *European Biophysics Journal* 30.1 (2001), pp. 75–81.

272 V. M. Dadarlat and R. D. Skeel. "Dual Role of Protein Phosphorylation in DNA Activator/coactivator Binding." *Biophysical Journal* 100.2 (2011), pp. 469–477.

273 T. Kiefhaber, A. Bachmann, and K. S. Jensen. "Dynamics and Mechanisms of Coupled Protein Folding and Binding Reactions." *Current Opinion in Structural Biology* 22.1 (2012), pp. 21–29.

274 K. Sugase, H. J. Dyson, and P. E. Wright. "Mechanism of Coupled Folding and Binding of an Intrinsically Disordered Protein." *Nature* 447.7147 (2007), pp. 1021–1025.

275 A. G. Turjanski, J. S. Gutkind, R. B. Best, and G. Hummer. "Binding-induced Folding of a Natively Unstructured Transcription Factor." *PLoS Computational Biology* 4.4 (2008), e1000060.

276 G. Bussi, D. Donadio, and M. Parrinello. "Canonical Sampling Through Velocity Rescaling." *Journal of Chemical Physics* 126.1 (2007), p. 014101.

277 H. J. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. Haak. "Molecular Dynamics With Coupling to an External Bath." *Journal of Chemical Physics* 81.8 (1984), pp. 3684–3690.

278 C. H. Bennett. "Efficient Estimation of Free Energy Differences From Monte Carlo Data." *Journal of Computational Physics* 22.2 (1976), pp. 245–268.

279 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. "Improved Side-chain Torsion Potentials for the Amber ff99SB Protein Force Field." *Proteins: Structure, Function, and Bioinformatics* 78.8 (2010), pp. 1950–1958.

280 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. "Comparison of Simple Potential Functions for Simulating Liquid Water." *Journal of Chemical Physics* 79.2 (1983), pp. 926–935.

281 N. Homeyer, A. H. Horn, H. Lanig, and H. Sticht. "AMBER Force-field Parameters for Phosphorylated Amino Acids in Different Protonation States: Phosphoserine, Phosphothreonine, Phosphotyrosine, and Phosphohistidine." *Journal of Molecular Modeling* 12.3 (2006), pp. 281–289.

282 M. Zacharias, T. Straatsma, and J. McCammon. "Separation-shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration." *Journal of Chemical Physics* 100.12 (1994), pp. 9025–9031.

283 T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren. "Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations." *Chemical Physics Letters* 222.6 (1994), pp. 529–539.

284 J. Anwar and D. M. Heyes. "Robust and Accurate Method for Free-energy Calculation of Charged Molecular Systems." *Journal of Chemical Physics* 122.22 (2005), p. 224117.

285 N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. "Electrostatics of Nanosystems: Application to Microtubules and the Ribosome." *Proceedings of the National Academy of Sciences of the United States of America* 98.18 (2001), pp. 10037–10041.

286 R. B. Best and G. Hummer. "Optimized Molecular Dynamics Force Fields Applied to the Helix–Coil Transition of Polypeptides." *Journal of Physical Chemistry B* 113.26 (2009), pp. 9004–9015.

287 E. J. Sorin and V. S. Pande. "Exploring the Helix–Coil Transition via All-atom Equilibrium Ensemble Simulations." *Biophysical Journal* 88.4 (2005), pp. 2472–2493.

288 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. "The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method." *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021.

289 A. Grossfield. "WHAM: The Weighted Histogram Analysis Method." *version 2.07* (), http://membrane.urmc.rochester.edu/content/wham, 2016/01/02.

290 S. Neumaier, A. Reiner, M. Büttner, B. Fierz, and T. Kiefhaber. "Testing the Diffusing Boundary Model for the Helix–Coil Transition in Peptides." *Proceedings of the National Academy of Sciences of the United States of America* 110.32 (2013), pp. 12905–12910.

291 S. Müller-Späth, A. Soranno, V. Hirschfeld, H. Hofmann, S. Rüegger, L. Reymond, D. Nettels, and B. Schuler. "Charge Interactions Can Dominate the Dimensions of Intrinsically Disordered Proteins." *Proceedings of the National Academy of Sciences of the United States of America* 107.33 (2010), pp. 14609–14614.

292 H. Matsuno, H. Furusawa, and Y. Okahata. "Kinetic Study of Phosphorylation-Dependent Complex Formation Between the Kinase-Inducible Domain (KID) of CREB and the KIX Domain of CBP on a Quartz Crystal Microbalance." *Chemistry A European Journal* 10.23 (2004), pp. 6172–6178.

293 **R. Bomblies**, M. P. Luitz, and M. Zacharias. "Molecular Dynamics Analysis of 4E-BP2 Protein Fold Stabilization Induced by Phosphorylation." *Journal of Physical Chemistry B* (2016), in press.

294 F.-X. Theillet, C. Smet-Nocca, S. Liokatis, R. Thongwichian, J. Kosten, M.-K. Yoon, R. W. Kriwacki, I. Landrieu, G. Lippens, and P. Selenko. "Cell Signaling, Post-translational Protein Modifications and NMR Spectroscopy." *Journal of Biomolecular NMR* 54.3 (2012), pp. 217–236.

295 G. A. Khoury, R. C. Baliban, and C. A. Floudas. "Proteome-wide Post-translational Modification Statistics: Frequency Analysis and Curation of the Swiss-prot Database." *Scientific Reports* 1 (2011), p. 90.

296 P. Cohen. "The Regulation of Protein Function by Multisite Phosphorylation–a 25 Year Update." *Trends in Biochemical Sciences* 25.12 (2000), pp. 596–601.

297 L. N. Johnson and R. J. Lewis. "Structural Basis for Control by Phosphorylation." *Chemical Reviews* 101.8 (2001), pp. 2209–2242.

298 E. S. Groban, A. Narayanan, and M. P. Jacobson. "Conformational Changes in Protein Loops and Helices Induced by Post-translational Phosphorylation." *PLoS Computational Biology* 2.4 (2006), e32.

299 A. Bah, R. M. Vernon, Z. Siddiqui, M. Krzeminski, R. Muhandiram, C. Zhao, N. Sonenberg, L. E. Kay, and J. D. Forman-Kay. "Folding of an Intrinsically Disordered Protein by Phosphorylation as a Regulatory Switch." *Nature* 519.7541 (2015), pp. 106–109.

300 J. L. Banko, F. Poulin, L. Hou, C. T. DeMaria, N. Sonenberg, and E. Klann. "The Translation Repressor 4E-BP2 Is Critical for eIF4F Complex Formation, Synaptic Plasticity, and Memory in the Hippocampus." *Journal of Neuroscience* 25.42 (2005), pp. 9581–9590.

301 N. Sonenberg and A. G. Hinnebusch. "Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets." *Cell* 136.4 (2009), pp. 731–745.

302 C. M. Fletcher, A. M. McGuire, A.-C. Gingras, H. Li, H. Matsuo, N. Sonenberg, and G. Wagner. "4E Binding Proteins Inhibit the Translation Factor eIF4E Without Folded Structure." *Biochemistry* 37.1 (1998), pp. 9–15.

303 P. Gosselin, N. Oulhen, M. Jam, J. Ronzca, P. Cormier, M. Czjzek, and B. Cosson. "The Translational Repressor 4E-BP Called to Order by eIF4E: New Structural Insights by SAXS." *Nucleic Acids Research* 39.8 (2011), pp. 3496–3503.

304 S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw. "Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States." *Journal of Physical Chemistry B* 119.16 (2015), pp. 5113–5123.

305 B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl. "GROMACS 4: Algorithms for Highly Efficient, Load-balanced, and Scalable Molecular Simulation." *Journal of Chemical Theory and Computation* 4.3 (2008), pp. 435–447.

306 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. "The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method." *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021.

307 R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, and R. C. Walker. "Routine Microsecond Molecular Dynamics Simulations With AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald." *Journal of Chemical Theory and Computation* 9.9 (2013), pp. 3878–3888.

308 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling. "Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters From ff99SB." *Journal of Chemical Theory and Computation* 11.8 (2015), pp. 3696–3713.

309 W. Humphrey, A. Dalke, and K. Schulten. "VMD: Visual Molecular Dynamics." *Journal of Molecular Graphics* 14.1 (1996), pp. 33–38.

310 D. R. Roe and T. E. Cheatham III. "PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data." *Journal of Chemical Theory and Computation* 9.7 (2013), pp. 3084–3095.

311 N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. "Electrostatics of Nanosystems: Application to Microtubules and the Ribosome." *Proceedings of the National Academy of Sciences of the United States of America* 98.18 (2001), pp. 10037–10041.

312 C. N. Schutz and A. Warshel. "What Are the Dielectric "Constants" of Proteins and How to Validate Electrostatic Models?" *Proteins: Structure, Function, and Bioinformatics* 44.4 (2001), pp. 400–417.

313 J. Morales, O. Mulner-Lorillon, B. Cosson, E. Morin, R. Bellé, C. A. Bradham, W. S. Beane, and P. Cormier. "Translational Control Genes in the Sea Urchin Genome." *Developments in Biologicals* 300.1 (2006), pp. 293–307.

314 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell Jr. "CHARMM36m: an improved force field for folded and intrinsically disordered proteins." *Nature Methods* (2016).