Studying the binding of phosphorylated peptides to proteins by molecular dynamics simulations

Dissertation zur Erlangung des Grades des Doktors der
Naturwissenschaften der Naturwissenschaftlich-Technischen
Fakultät der Universität des Saarlandes

von
Nicolas KÜNZEL

Saarbrücken, September 2021

Tag des Kolloquiums	24.01.2022
Dekan	Prof. Dr. Jörn Walter
Vorsitz	Prof. Dr. Uli Müller
Erstgutachter	Prof. Dr. Volkhard Helms
Zweitgutachter	Prof. Dr. Jochen Hub
Akademischer Mitarbeiter	Dr. Mark Lommel

Abstract

The knowledge about protein-protein interactions is rising steadily. Protein domains and signaling proteins in particular are becoming increasingly important in the search for new drugs to combat many diseases, as they are commonly found in central positions in protein interaction cascades. They often form complexes with specific types of peptides which is why it is interesting to study the interaction properties as well as the way in which the binding can be influenced. For the investigation of such complexes, computational methods can be used to get a deeper understanding of the mechanistical binding properties.

In this thesis, I present our approaches on understanding the interactions of 14-3-3 proteins and PDZ domains with flexible peptides, intrinsically disordered regions of the partner proteins. Thereby, the main focus is on the influence of phosphorylation of serine residues in the peptides and how this effects the binding. Thus, I illustrate the results of various plain molecular dynamics simulations to study the mechanistic details of the interactions. Furthermore, the results of alchemical simulations for the calculation of relative binding free energies and parallel cascade selection molecular dynamics (PaCS-MD) simulations for the calculation of absolute binding free energies are outlined.

Finally, I present a project where we studied the competitive binding between a peptide and an inhibitor to the same binding pocket of a PDZ domain using PaCS-MD simulations.

Zusammenfassung

Das Wissen über Protein-Protein-Wechselwirkungen nimmt stetig zu. Besonders Proteindomänen sowie Signalproteine gewinnen bei der Suche nach neuen Medikamenten zur Bekämpfung diverser Krankheiten zunehmend an Bedeutung, da sie häufig im Zentrum von Proteinwechselwirkungskaskaden stehen. Da diese oft Komplexe mit bestimmten Peptiden bilden, ist es interessant, Interaktionseigenschaften und Möglichkeiten zur Beeinflussung der Bindung zu untersuchen. Dabei können computergestützte Methoden eingesetzt werden, um ein tieferes Verständnis der mechanistischen Bindungseigenschaften zu erlangen.

In dieser Arbeit stelle ich unsere Ansätze zum Verständnis der Wechselwirkungen von 14-3-3-Proteinen und PDZ-Domänen mit flexiblen Peptiden vor. Hierbei haben wir vor allem den Einfluss der Phosphorylierung von Serinresten in den Peptiden auf die Bindungen untersucht. Es werden die Ergebnisse verschiedener Molekulardynamik (MD)-Simulationen vorgestellt, in denen die mechanistischen Details der Wechselwirkungen untersucht wurden. Weiterhin werden die Ergebnisse alchemistischer Simulationen zur Berechnung relativer und sogenannter "parallel cascade selection molecular dynamics" (PaCS-MD) Simulationen zur Berechnung absoluter freier Bindungsenergien dargestellt.

Schließlich stelle ich ein Projekt vor, in dem wir die kompetitive Bindung zwischen einem Peptid und einem Inhibitor an dieselbe Bindungstasche einer PDZ-Domäne mit Hilfe von PaCS-MD-Simulationen untersucht haben.

Acknowledgements

The last three and something years were a very defining time for my life. I want to use this opportunity to thank every person who helped and supported me during this time and also during my whole life. Without this support I would not be at this point right now.

First of all, I would like to thank Prof. Dr. Volkhard Helms for the opportunity to do my doctoral research in his group. Of course, I would also like to thank for his support and mentoring and his confidence in my abilities to carry out my projects independently.

I also want to thank Prof. Dr. Jochen Hub for taking the time to review my thesis and all my collaborators and coworkers from the different projects I participated in, namely Dr. Simone Trautmann, Prof. Dr. Barbara Niemeyer, Dr. Mona Knapp, Dr. Claudia Fecher-Trost, Prof. Dr. Ivan Bogeski, Sabrina Capello, Andreas Denger, Dr. Björn Becker and Dr. Johanna Dudek.

Additionally, I want to thank my former and current colleagues, especially Dr. Thorsten Will and PD Dr. Michael Hutter, who helped me on getting aboard the group and my projects and spend a lot of time in the Mensa with me. A special thanks goes to Kerstin Gronow-Pudelek, our great secretary, who was always very supportive and helpful for all kinds of organizational questions and issues.

I want to thank my whole family for their support in all the good and bad times during my whole life. Especially, of course, my parents, who have always been by my side and who have played a big part in making me the person I am today. Also my brother, Malte, who always smoothed the waters when the mood was not so good. All of you always unconditionally supported me, often helped me to overcome motivational holes and gave me useful advice, when I needed it. Thank you! I deeply love all of you.

Finally, I want to thank the most important person in my life, Marie. Without you, I would probably have abandoned my doctorate several times in between and would likely have often despaired. Your love and support was the most important anchor during this time. I am incredibly happy to have you by my side. I love you!

Contents

List of Figures

List of Tables

Glossary

ATP adenosine triphosphate

BAR Bennett acceptance ratio

CGenFF CHARMM General Force Field [1, 2]
CHARMM Chemistry at HARvard Macromolecular Mechanics program [3]
COM center of mass

FF force field

GROMACS Groningen Machine for Chemical Simulations [4]

MC Monte-Carlo
MD molecular dynamics
MM molecular mechanics
MP2 second-order Møller-Plesset perturbation theory [5]
MSM Markov state model

PaCS parallel cascade selection
PBC periodic boundary condition
PBM PDZ binding motif
PCA principal component analysis
PDB Protein Data Bank (www.rcsb.org) [6, 7]
PDZ PSD-95/Discs-large/ZO-1 (domain)
PME particle-mesh Ewald [8, 9]
PMF potential of mean force
PTM post-translational modification

RMSD root mean squared deviation

SASA solvent accessible surface area
SER serine
SPC simple point charge (water model) [10]

SUZ nonsteroidal anti-inflammatory drug sulindac (trade name: Clinoril), IUPAC name: 2-[(3Z)-6-fluoro-2-methyl-3-[(4-methylsulfinylphenyl)methylidene]inden-1-yl]acetic acid [11]

tIC time-lagged independent component

tICA time-lagged independent component analysis [12–15]

TPT transition path theory [16]

Chapter 1

Introduction

1.1 Motivation

Without proteins and their interactions, affected and altered by various external and internal influences, life as we know it would not be possible. It is estimated, that mammals have between $10^9 - 10^10$ proteins per cell [17] making up around 60 % of the cells dry weight [18]. Moreover, proteins are involved in nearly every cellular process due to their interactions with other proteins, carbohydrates, lipids, RNA or DNA.

Today, most of our understanding about protein-protein interactions and their stability is derived from three-dimensional atomistic structures of protein-protein complexes (see among others Jones and Thornton [19] and Marsh and Teichmann [20]). Thus, to study protein-protein, protein-peptide or protein-small molecule interactions information about the secondary, i. e. the shape of protein segments, and tertiary structure, i e. the shape of the full proteins, is needed. Usually these structures are experimentally observed and characterized by structural biology groups using nuclear magnetic resonance spectroscopy or X-ray crystallography that allow to obtain three-dimensional structures of proteins in high resolution under specific conditions. These processes are very time demanding and costly. However, computer science can help with this structure determination as was just recently demonstrated by the group developing the AlphaFold software for predicting the three-dimensional structure of proteins very accurately [21] or since many years by the group of Vijay Pande with their Folding@home software [22, 23] even though experimental validation is always wanted. Similar computational support for the understanding of biological processes can be given by the help of the molecular dynamics (MD) or Monte-Carlo methods that use the determined protein structures as a starting point. Usual experiments measure only the states of a system but the process in between stays fuzzy. This is where the computational methods come into play. Simply put, they can take the initial state of a biological system and propagate the system until it ends up in the final state and thus it is possible to analyze the intermediary process. Until today, computational studies trying to characterize protein-protein or protein-peptide interactions using MD simulations in explicit solvent are very scarce. Though, a few examples exist. For example the binding of a phosphopeptide to an SH2 domain [24], a peptide binding to a PDZ domain [25] or a study characterizing the binding free energy between a SH2 domain and a peptide [26]. The existing studies are promising that we are still just in the beginning of this field because advances are made every day, e. g. Markov state models (MSMs) come into play for analysing MD trajectories of protein-peptide interactions [27, 28].

Protein-protein interactions are affected by various effects, e. g. post-translational modifications (PTMs). These highly increase the amount of protein states resulting in even more possible interactions. PTMs can also strongly influence interactions,

e. g. they can be interaction switches between the binding to 14-3-3 proteins or PDZ domains [29–31]. However, the influence and specific effects of peptide phosphorylation, one of the most common post-translational modification, on the binding to proteins has not been targeted in much detail in many experimental or computational studies so far.

Thus, in this work, the main focus was put on the binding of peptides to proteins in general but mostly how this binding is influenced by phosphorylation of serine residues in the peptide and how this can be studied computationally. Additionally, we were interested in how the competitive binding of peptides and their inhibitors to proteins can be simulated. We wanted to study these using MD simulations, enhanced sampling and free energy methods.

1.2 Overview and Objectives of this thesis

This thesis mainly focuses on the binding of peptides to PDZ domains and 14-3-3 proteins and how this binding is affected by phosphorylation of serine residues in the peptides. Additionally, the competition between the binding of a peptide and a small molecule inhibitor to a PDZ domain was studied. In general, we wanted to get a deeper understanding of (mostly disordered) peptide binding to PDZ and 14-3-3 domains and how this can be influenced through mutation, post-translational modifications and inhibition.

During the course of my doctoral studies I was also part of several additional projects which were not directly related to protein-peptide binding but focused on other properties and task of proteins in mammals. They allowed me to gain broader insight into the field of computational biophysics and proteomics, helping to understand the greater context of processes in our cells. Two projects focused on the adsorption of salivary proteins on various materials including enamel. The aim was to understand the differences between the protein concentrations in saliva, on enamel and materials which could be used for dental prosthesis in order to gain insight into how different materials behave in the mouth and what protein properties lead to adsorption of proteins to these materials. Here, I increased my knowledge on statistical methods and the analysis of proteomics data and got deeper insight into experimental methods in proteomics. In another project the NK cell cytotoxicity for melanoma cells was studied. The focus was on finding markers for and pathways involved in melanoma killing and developing algorithms to predict NK-cell-mediated melanoma killing for specific cell lines. In this project, I further developed my knowledge about statistics and learned a lot about various machine learning techniques. Further I got insight into possible cancer therapies and ways to develop possible new therapeutic techniques. The last project I participated in focussed on the interaction between the alternative spliced form of the STIM1 protein (STIM1A) and the hOrai1 protein and the specific functions of this spliced variant. In this project I dived deeper into molecular docking and homology modeling.

1.2.1 First author publications/projects and additional projects related to doctoral studies

N. Künzel, V. Helms, **"How peptide phosphorylation affects its interaction with 14-3-3η domains"**, PROTEINS: Structure, Function, and Bioinformatics, Accepted Author Manuscript, 2021, `https://doi.org/10.1002/prot.26224` [32].
Abstract:

Members of the 14-3-3 domain family have important functions as adapter domains. Via an amphipathic groove on their protein surface they typically bind to disordered C-terminals of other proteins. Importantly, binding partners of 14-3-3 domains usually contain a phosphorylated serine or threonine residue at their binding interface and possess one of three different sequence motifs. Binding of the respective unphosphorylated versions of the peptides is typically strongly disfavored. There is a wealth of structural and thermodynamic data available for the phosphorylated forms but not for the unphosphorylated forms as the binding affinities seem to be too weak to be measurable experimentally. Here, we characterized the mechanistic details that govern the preference for the binding of phosphorylated peptides to 14-3-3η domains by means of molecular dynamics simulations. We found that the phosphate group is ideally coordinated in the binding pocket whereas the respective unphosphorylated side-chain counterpart is not. Thus, the binding preference results from the tight coordination of the phosphorylated residue at the center of the binding interface. Furthermore, MD simulations of 14-3-3η dimers showed a preference for the simultaneous binding of two phosphorylated peptides in agreement with their experimentally observed cooperativity.

My contribution: I chose the systems to study and the simulations to perform, performed the simulations, analyzed the data and wrote the first draft of the manuscript (see chapter 3 for preprint version of manuscript).

N. Künzel, V. Helms, **"How peptides bind to PDZ domains"**, in preparation
Abstract:
The highly abundant PDZ domains, protein modules involved in many protein-protein interactions, play an important role in the assembly of supramolecular complexes and signal transduction. They usually bind to the C-terminals of other proteins via three main sequence motifs. Interactions are often controlled by phosphorylation of the peptide at the serine or threonine residue of motif 1. Here, we studied the mechanistic details of motif 1 peptides with and without a phosphate group attached to the central serine residue, as well as a peptide were serine is mutated to glutamic acid, binding to the hPTP1E PDZ2 domain and the MAGI1 PDZ1 domain using molecular dynamics simulations. We identified general binding features for peptides binding to PDZ domains. These are backbone-backbone interactions with the beta strand in the binding pocket and specifically very stable interactions involving the backbone of the residue at position -2 from the C terminus (often a phosphorylatable residue is found at this position). The binding is further stabilized by interactions of the hydrophobic C-terminal residue of the peptide with the surrounding residues of the PDZ domain. Additionally, we observed that specific residues in the PDZ domains have a strong influence on the binding of specifically phosphorylated peptides and other peptides that contain a negative charge at position -2 from the C terminus to these domains. These residues are ARG79 and LYS38 in the hPTP1E PDZ2 domain and equivalently GLN85 and LYS44 in the MAGI1 PDZ1 domain if mutated to arginine, as well as LYS72 in the hPTP1E PDZ2 domain.
My contribution: I chose the systems to study and the simulations to perform, performed the simulations, analyzed the data and wrote the first draft of the manuscript (see chapter 4).

N. Künzel, V. Helms, **"How to study competitive binding computationally"**, work

in progress
Abstract:

My contribution: I chose the systems to study and the simulations to perform, performed the simulations, analyzed the data and wrote the first draft of a potential manuscript (see chapter 5).

1.2.2 Coauthor publications and additional projects during doctoral studies

S. Trautmann, N. Künzel, C. Fecher-Trost, A. Barghash, P. Schalkowsky, J. Dudek, J. Delius, V. Helms, M. Hannig, **"Deep Proteomic Insights into the Individual Short-Term Pellicle Formation on Enamel—An In Situ Pilot Study."** *Prot. Clin. Appl.* 2020, 14, 1900090. `https://doi.org/10.1002/prca.201900090` [33]
Abstract:
Purpose
Dental pellicle formation starts instantaneously after oral hygiene due to the adsorption of salivary proteins to all orally exposed surfaces. The pellicle acts as a physiological mediator, protects the tooth surface from mechanical damages and reduces acid-induced enamel demineralization. The aim of this pilot study is to identify and characterize individual proteomic profiles of the initial pellicle formed on dental enamel and to compare the profiles with the corresponding saliva to analyze specific adsorption patterns occurring during pellicle formation.
Experimental Design
The 3-min pellicle of five subjects formed in situ on bovine enamel is eluted chemically and analyzed separately by nano-mass spectrometry. The analysis of the corresponding saliva is conducted in parallel.
Results
Up to 498 pellicle proteins and up to 1032 salivary proteins are identified on an individual level. Comparison of the salivary and pellicle protein profiles demonstrates the pellicle formation to be highly individual. Nineteen proteins are significantly enriched in the 3-min pellicle of all subjects and 22 proteins are significantly depleted indicating that pellicle formation relies on selective adsorption.
Conclusions and Clinical Relevance
The short-term enamel pellicle is composed of several hundreds of adsorbed salivary proteins and reveals a highly individual proteomic profile.

My contribution: I performed qualitative and quantitative proteomics analysis including fold changes, molecular weight distributions, isoelectric point distributions, GO terms, molecular functions and GO-term enrichment. I also wrote the first drafts of the respective method and results sections and created most of the figures in the manuscript.

S. Trautmann, N. Künzel, C. Fecher-Trost, A. Barghash, J. Dudek, V. Flockerzi, V. Helms, and M. Hannig, **"Is the proteomic composition of the salivary pellicle dependent on the substrate material below?"**, in preparation
Abstract:
The use of dental restorative materials is a routine task in clinical dentistry. Upon exposure to the oral cavity, continuous adsorption of salivary proteins and other macromolecules to all surfaces occurs, representing the process of dental biofilm

formation. Each substrate material has different physico-chemical properties and this may influence the composition of the initial biofilm, termed pellicle. This study aimed at characterizing the individual proteomic composition of the 3-min pellicle formed on bovine enamel and six restorative materials. Substrate material-specific adsorption patterns were analyzed by comparing the proteomic profiles of the 3-min pellicle with the corresponding saliva. The combination of chemical elution and nano-LC-HR-MS/MS resulted in the identification of 1348 different pellicle proteins, out of which 187 to 686 proteins were present in individual 3-min pellicles. Proteomic profiles were analyzed in terms of molecular weights, isoelectric points, and molecular functions of the identified proteins. Unexpectedly, this yielded quite similar distribution patterns independent of the substrate materials. Furthermore, overall similar fold changes were obtained for the major part of commonly enriched or depleted proteins in the 3-min pellicles. Taken together, the current results point to a minor important role of the substrate material on the proteomic composition of the 3-min pellicle.

My contribution: I performed qualitative and quantitative proteomics analysis including fold changes, molecular weight distributions, isoelectric point distributions, GO terms, molecular functions and GO-term enrichment. I also wrote the first drafts of the respective method and results sections and created most of the figures in the manuscript.

S. Cappello, H.-M. Sung, C. Ickes, C. Gibhardt, A. Vultur, H. Bhat, Z. Hu, P. Brafford, A. Denger, I. Stejerean-Todoran, R.-M. Köhn, V. Lorenz, N. Künzel, G. Salinas, H. Stanisz, T. Legler, P. Rehling, M. P. Schön, K. S. Lang, V. Helms, M. Herlyn, M. Hoth, C. Kummerow, I. Bogeski, **NK cell cytotoxicity and protein microarrays predict efficacy of melanoma immunotherapies**, Cancer Research , Accepted Author Manuscript, 2021

Abstract:

Notwithstanding the impressive advances in melanoma-directed immunotherapies, resistance is common and many patients still succumb to the metastatic disease. In this context, natural killer (NK)-cells, although side-lined in the recent development of melanoma immunotherapy, could provide therapeutic benefits in the future. To identify molecular determinants of NK-cell-mediated melanoma killing (*NKmK*), we quantified NK-cell cytotoxicity against a panel of genetically-diverse melanoma cell lines and observed a highly heterogeneous susceptibility. Melanoma cell protein microarrays revealed a correlation between protein abundance/activation and *NKmK*. A "protein-killing-signature" identified metabolic factors as essential regulators of *NKmK*. Using 2D and 3D killing assays and melanoma xenografts, we demonstrated that the PI3K/Akt/mTOR signaling-axis controls *NKmK* via expressional regulation of NK cell-relevant surface proteins. Moreover, we developed algorithms to predict *NKmK* of additional melanoma cell lines and the response of melanoma patients to anti-PD-1 checkpoint therapy. Our findings identify novel NK-cell-related prognostic biomarkers and might thus contribute to improved and personalized melanoma-directed immunotherapies.

My contribution: I assisted in analyzing the RPPA dataset and finding proteins strongly correlated with NK-cell killing using machine learning techniques to predict whether NK-cell killing will work well or not for specific cell lines.

M. L. Knapp, D. Alansary, K. Förderer, F. Sommer, D. Zimmer, Y. Schwarz, N. Künzel, A. Kless, K. Machaca, V. Helms, T. Mühlhaus, M. Schroda, A. Lis, B. A. Niemeyer, **"A longer isoform of Stim1 is a negative SOCE regulator but increases cAMP modulated NFAT signaling"**, submitted, preprint available [34]

Abstract:

Alternative splicing is a potent modifier of protein function. Stromal interaction molecule 1 (Stim1) is the essential activator of store-operated Ca^{2+} entry (SOCE) necessary to trigger activation of transcription factors such as NFAT. Here, we characterize Stim1A, a splice variant with an additional 31 amino acid domain inserted in frame within its cytosolic domain. Prominent expression of *exon A* is found in astrocytes, heart, kidney and testes. Full length Stim1A functions as a dominant-negative regulator of SOCE and I_{CRAC}, facilitating fast calcium dependent inactivation (FCDI). Docking analysis guided mutational analysis shows that STIM1A destabilizes gating of Orai1 in a sequence-specific manner with downregulation or absence of native Stim1A resulting in increased SOCE. Despite reducing SOCE, Stim1A leads to increased NFAT translocation. Differential proteomics revealed interference of Stim1A with the cAMP-SOCE crosstalk by altered modulation of phosphodiesterase (PDE8B), resulting in differential cAMP modulated translocation of NFAT. Our study uncovers cell type specific splicing of Stim1 as a potent means to regulate $G_{\alpha s/\alpha i}$ input on the cAMP-SOCE crosstalk.

My contribution: I performed homology modeling and molecular docking with various settings, analyzed the docked structures in terms of interesting amino acid interactions and created the respective figures.

1.3 Outline

After introducing the biological and theoretical background for this thesis in chapter 2, the three first-author projects I have worked on during my doctoral research will be presented. The first two projects are presented as finished manuscripts in the form that was submitted to scientific journals while the results from the third project will be submitted in the future. If needed, an addendum is given for a project. For the first project some additional analyses are included that did not end up in the manuscript.

Within chapter 3, the first-author project that will be presented, focused on the influence of serine phosphorylation on the binding of peptides (RSRST[S/pS]TP[NV] and RLYH[S/pS]LP) to 14-3-3 proteins, specifically the 14-3-3η protein. The main methodology for the alchemical simulations and the respective charge corrections is shown which was also used in the following project. The latter will be presented thoroughly in chapter 4 and focuses on the binding of differently charged peptides (EQV[S/E/pS]AV) to PDZ domains and on the influence of mutations of positively charged amino acids, especially a specific arginine, to uncharged ones in the binding pocket of the studied PDZ domains, hPTP1E PDZ2 and MAGI1 PDZ1. Moreover, also the methodology for parallel cascade selection molecular dynamics (PaCS-MD) simulations is presented. This is also used in the third project, that will be described in chapter 5. The aim of this project was to study the competitive binding of a small molecule inhibitor and a peptide to a PDZ domain.

In the end, chapter 6 gives concluding remarks and a general outlook on the research topic and beyond.

Chapter 2

Background

In order to present the theoretical background of the thesis, this chapter provides the biological and theoretical background for the systems studied and methods used in this thesis. Furthermore, it also extends on additional methods that could have also been used to study binding free energies so that the methods used can be easily classified. Of course, the focus is mainly placed on the background which is needed to understand the projects of this thesis and does not claim to give an overall coverage of the full topics.

In this chapter parts will also be part of a chapter I wrote together with Volkhard Helms for a book that will be published by Wiley-VCH under the title "Protein Interaction. The Molecular Basis of Interactomics". My contribution to this chapter was to collect the content and to write the first draft of the chapter. The parts that include content that will likely also be in the book chapter are parts of sec. 2.2 including sec. 2.2.1 and parts of sec. 2.5, specifically sec. 2.5.2 with subsections, sec. 2.5.3 with subsections, sec. 2.5.4 with subsections and sec. 2.5.5.

2.1 Biology

This section serves to give an introduction into the general biological concepts and systems that have been studied in this thesis. First of all, post-translational modifications are introduced whereby a special focus is placed on phosphorylation. Afterwards, the two classes of proteins or protein domains studied in this thesis are introduced, namely 14-3-3 proteins and PDZ domains.

2.1.1 Post-translational modifications

Post-translational modifications (PTMs) are chemical modifications due to covalent additions of functional groups or proteolytic cleavage of a polypeptide chain to allow forming of a mature protein [35]. The name is most often associated with the former ones and the focus in this thesis will also be put on these. The covalent modifications allow for interactions and therefore functionalities of proteins which are not possible when only considering the genetically stored information. The combination of multiple PTMs leads to an enormous amount of potential molecular states, especially, because these modifications can happen on a developmental or physiological time scale and not only on an evolutionary time scale like genetic evolution [36]. PTMs greatly increase the amount of possible specific proteins and protein states. While only 20000-25000 genes exist in the human genome [35, 37] over 1 million human proteins or protein states are known to exist because single genes encode multiple proteins, alternative splicing takes place and most importantly due to the sheer amount of possible PTMs [35, 38], carried out by enzymes such as kinases, ligases, phosphatases, etc. that can either add or remove functional groups on amino acid side chains or peptide linkages [35]. Multiple PTMs can modify a protein at the same time on different positions or on the same site at different times [39].

PTMs can either be stable, i. e. they stay continuously on a protein, or more volatile. The former ones, examples are lipidation, glycosylation and disulfide bridge formation, can help in protein maturation and folding. The latter ones, a prominent example is phosphorylation, often play a role in intracellular signaling [39] that is usually processed by modular protein domains or special proteins which recognize specific types of PTMs on specific residues, like PDZ domains (see sec. 2.1.3) or 14-3-3 proteins (see sec. 2.1.2). Due to their important role in many crucial functions of proteins and cellular processes, PTMs and their dysregulation are associated with various diseases such as cancer [40–45], cardiovascular diseases [40, 46–48] and diabetes [49–51].

Phosphorylation

Phosphorylation of proteins, first detected as phosphoserine in 1932 [52] and as enzymatic process in 1954 [53], is the most commonly observed PTM [45, 54] (an estimated amount of 30 % of all cellular proteins are phosphorylated on at least one residue [55–57]) and the main study object of this thesis. Phosphorylation is associated with various processes in biological cells, especially signaling and regulation [58], e. g. through modifications of PDZ or 14-3-3 interactions (see sec. 2.1.3 and 2.1.2). Phosphorylation sites are often located at binding interfaces, are more likely to be evolutionary conserved than other residues at interfaces and can strongly influence the binding energies of protein-protein interactions [59]. Intrinsically disordered proteins and flexible protein regions show tendencies to be phosphorylated and phosphorylation may cause strong structural changes in these regions, like disorder-to-order and order-to-disorder transitions [59–63]. Phosphorylation and dephosphorylation are regulated by specific classes of enzymes, namely protein kinases and phosphates, respectively. The former are encoded by around 2 % of the human genome [64, 65] reflecting the importance of these processes. Nine out of the 20 natural amino acids in proteins can be phosphorylated by kinases to form phosphate esters usually using adenosine triphosphate (ATP) (and sometimes guanosine triphosphate and phosphoenolpyruvate) as a phosphate donor [66]. These amino acids are serine (SER), threonine and tyrosine (most commonly in eukaryotes), as well as histidine, aspartic acid and glutamic acid (most commonly in bacteria [67–69]) and also arginine, cysteine and lysine [66]. Phosphate, PO_4^{3-}, is a molecule that has a high solubility in water, forms a large hydrated ionic shell and is found in a dianionic form at pH 7 due to its three pK_a values (2.2, 7.2 (which is 5.8 for an ester) and 12.4) [66]. For example, for phosphoserine the pK_a values are 2.19, 5.78 and 9.85 [70, 71].

Phospho-amino acids do not resemble any natural amino acid and can therefore increase the chemical diversity in proteins. The phosphate group, usually dianionic at physiological pH, differs from the negatively charged amino acids, aspartate and glutamate, which have a smaller hydrated shell and carry only one negative charge [66]. Therefore, these two are not ideal phosphomimetics on their own. Additionally, the phosphate group can engage in stronger and more stable hydrogen bonds and salt bridges especially with arginine than aspartate and glutamate [66, 72]. Hydrogen bond strengths are similar when comparing the monoanionic and dianionic states of the phosphate but the protonation state of the phosphate can still be a critical parameter [72].

In the light of this background, the focus of this thesis is put on the structural and thermodynamic implication of serine phosphorylation in peptides binding to PDZ and 14-3-3 domains.

2.1.2 14-3-3 proteins

The first objects of interest for this thesis were 14-3-3 proteins, since they are known to predominantly bind phosphorylated sequence motifs of their partner proteins. 14-3-3 proteins are a highly conserved protein family with seven isoforms in mammals [73]. These isoforms exhibit high structural and sequence similarity (RMS deviations between 0.7 and 1.8 Å) [74]. They are named α/β, γ, ϵ, η, ζ/δ, θ, and σ. A search on Google Scholar in July 2021 listed around 705 articles published since 2018 that have "14-3-3" in their title, evidence on the high relevance of 14-3-3 proteins in various fields of research. 14-3-3 proteins are associated with apoptosis, adhesion, cellular proliferation, differentiation and survival, cell cycle regulation and signal transduction pathways [73, 75] and are abundant in nearly all eukaryotic cells. They also play a role in cancer progression and are associated with various cancer types [76–79], such as breast [80, 81], lung [82] and gastric [83, 84] cancer. One example is their involvement in the p53 pathway which is a key tumor suppressor mechanism. They interact with various proteins in that pathway, thereby regulating wild-type p53 activity and they could be possible drug targets for cancer treatment [85]. Furthermore, 14-3-3 proteins are also linked to age-related neurodegenerative diseases like Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis and more [86–89]. Additionally, 14-3-3 proteins play a role in viral infection and replication [90, 91], e. g. they can interact with the Hepatitis C virus core protein [92] and the severe acute respiratory syndrome coronavirus (SARS-CoV) nucleocapsid protein by binding to a phosphorylation site [93]. Only recently, it was shown that also the SARS-CoV-2 nucleocapsid protein (N) binds to all seven human 14-3-3 isoforms, strictly depending on phosphorylation [94] and playing a role in the control and inhibition of the replication, transcription and packaging of the SARS-CoV-2 genome [95].

In general, 14-3-3 proteins exist as homo- and heterodimers [32, 97, 98], that are stably folded and stabilized because of hydrophobic interactions and salt bridges [32, 99, 100] (an example is given in fig. 2.1). Nine α-helices form an amphipathic groove and build up each monomer. This groove, where lysine and arginine residues form a basic cluster mediating the interaction with 14-3-3 binding partners [32, 101–103], is the predominant binding interface between target proteins and 14-3-3 [32, 103]. Moreover, they mainly interact with phosphorylated disordered regions of their partner proteins. In general, the most important binding motifs can be classified as motif 1(R[S/F/Y/W]XpSXP), motif 2 (RX[S/Y/FW/T/Q/A/D]Xp(S/T)X[P/L/M]) [104–106] and motif 3 (RXXp(S/T)XX-COOH) [105], but 14-3-3 proteins also interact with a few unphosphorylated peptides [101, 103, 107], e. g. the Carbohydrate-response Element-binding Protein (ChREBP) [108, 109] or the exoenzyme S (ExoS) [110].

During the course of my doctoral research I studied the binding of phosphorylated peptides and their unphosphorylated counterparts to 14-3-3 proteins to get a deeper understanding of the specificity of 14-3-3 proteins binding phosphorylated peptides. Within this domain family, the 14-3-3η domain was selected as the main system of interest because it has been shown to interact with various peptides for which the binding affinities have been characterized in experiments. Especially one

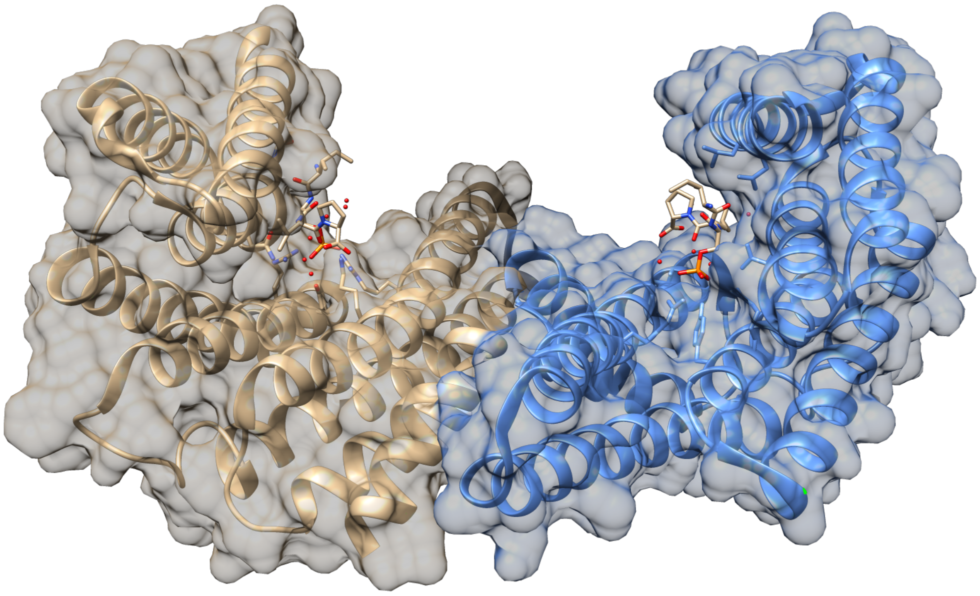


FIGURE 2.1: 14-3-3η homodimer binding a phosphorylated peptide (PDB-ID: 2C63 [96]). The protein is shown in ribbon style, the two peptides in atomistic stick mode.

study by Yaffe et al. [104] has to be mentioned here, because the binding of a phosphorylated peptide and the unphosphorylated counterpart to 14-3-3η was extensively studied. Thus, that study provided us with free energy values as references for our simulations. The 14-3-3η domain is encoded by the gene YWHAH, has a length of 246 amino acids. Furthermore, it is linked among others to rheumatoid arthritis [111–115] and could be a potential therapeutic target for cancers due to its role in mitotic progression [116].

2.1.3 PDZ domains

The second type of systems which were studied in this thesis are PDZ (PSD-95/Discs-large/ZO-1) [117] domains. These are a family of highly abundant modules in proteins that exist in various species and various proteins [118–121]. They are involved in many protein-protein interactions and play an important role in several biological processes, such as cell cycle, signal transduction, the assembly of supramolecular complexes and metabolism [120, 122–127]. A curated list of PDZ domains recently listed 272 different human PDZ domains in 154 human proteins [128]. In July 2021, a search in the Protein Data Bank [6] led to a list of 1961 structures that included PDZ domains and a search on Google Scholar listed around 3050 entries that included “PDZ” in their title, showing the broad interest in and importance of this domain family. Since they often have crucial roles in encounter pathways of protein-protein interaction, PDZ domains are a common target of viruses like adenoviruses, influenza viruses or human papillomaviruses [129–135]. Only recently it was shown that a protein-binding domain in the Envelope protein (E) of SARS-CoV-2 shows stronger binding to the PALS1 PDZ domain than the respective domain in E of SARS-CoV. This could be a critical step in triggering viral infection [135], since it was suggested before that the interaction between PALS1 and E of SARS-CoV plays an important role in its virulence [135, 136]. Besides, the E protein of SARS-CoV-2 was also shown to interact with the second PDZ-domain of the host tight junction protein ZO1 [137]. Additionally, Caillet-Saguy et al. found significant interactions with dissociation constants in the range of 3 to 82 µmol for sixteen human PDZ binders of the SARS-CoV-2 proteins E, 3A and N whereby six of them are also interacting with SARS-CoV proteins, including PTPN13 (also known as hPTP1E), and three were specific to SARS-CoV-2 E protein [138]. Furthermore, PDZ domains are associated with neurological diseases like Alzheimer's disease and Parkinson's disease, autism spectrum disorder and more [127, 139–142]. PDZ domain-containing proteins are also strongly involved in different stages and forms of cancer, i. e. from the formation of tumors to metastasis, mostly through their functions in signaling pathways [127, 130, 134, 143–146].

PDZ domains usually bind to their interactions partners via the very C-terminal ends of these proteins. There exist three main PDZ binding motifs (PBMs) that are commonly found. These are X[S/T]XΦ-COOH (motif I), XΦXΦ-COOH (motif II) and X[E/D]XΦ-COOH (motif III) where X can be any residue and ϕ means a hydrophobic residue [29, 121, 147, 148]. These interactions with motif I peptides can be controlled by phosphorylation at the serine or threonine residue, e. g. it was shown that phosphorylation of the C-terminal of PRMT5 modulates an interaction switch between a 14-3-3 and PDZ domain [29]. The peptides usually form complexes with a binding pocket of the PDZ domain formed by a beta strand and an alpha helix, antiparallel to the former.

In this thesis we will study the binding of different peptides to three PDZ domains. The first one is the PDZ2 domain of the human Tyrosine-protein phosphatase non-receptor type 13 (PTPL1, PTP-BAS, hPTP1E), coded by the gene PTPN13. This protein was first identified in 1994 under the name PTP-BAS [149] and shortly after renamed to hPTP1E [150]. The PDZ domains of hPTP1E are associated with carcinogenesis and/or cytoskeleton organization and cell migration [151], likely playing a role in several steps of tumor progression [152]. Moreover, the PDZ2 domain interacts with Fas [152, 153], RIL [152, 154], the tumor suppressor PTEN [155] and more [152]. Thus, we chose this PDZ domain because a publication reported experimental binding constants for unphosphorylated and phosphorylated peptides [156] which is seldomly found but is needed as reference when studying the influence of phosphorylation on the binding of peptides.

The second domain that will be studied is the PDZ1 domain of the human Membrane Associated Guanylate Kinase, WW And PDZ Domain Containing 1 (MAGI1) protein, first discovered in 1997 [157] and coded by the gene MAGI1. It is present in the cytoplasm and involved in cell-to-cell contacts [158–160]. It interacts, for example, with the tumor suppressor PTEN [161] and is a potential tumor suppressor for subtypes of breast cancer [162]. For the purpose of this thesis, this PDZ domain was chosen because using the BioGRID interaction database v3.5 [163] and the STRING database [164] we identified that it also binds the same EQVSAV peptide that binds to hPTP1E.

The third domain of interest for this thesis is the PDZ domain of the mouse (*mus musculus*) Dishevelled-1 (mDvl1) protein. Interacting with the membrane-bound receptor Frizzled in the Wnt signaling pathway [165], the Dvl1 PDZ domain was proposed as a drug target [166–169]. Moreover, Dvl proteins are associated with various cancer types [169–174]. We chose this PDZ domain in order to study competitive binding because a publication by Lee et al. reported experimental binding and inhibition constants for the peptide SGSLKLMTTV-COOH and the inhibitor Sulindac [175] and this system is therefore well suited for such a study.

2.2 Thermodynamic Ensembles and Free Energy

This section gives an introduction into the concept of thermodynamic ensembles and their connected state functions, the free energies. The focus is placed on the isothermal-isobaric ensemble in which most biological experiments are performed. Afterwards the Jarzynski equality is explained that is a special equation allowing to calculate equilibrium free energies from non-equilibrium work values which is later on used when analyzing the performed alchemical simulations.

The field of statistical mechanics links the classical microscopic states of a system to the macroscopic observables which can be measured in experiments. These can be thermodynamic, structural and dynamical properties. A key concept of statistical mechanics are thermodynamic ensembles [176] that characterize the probability distribution for all possible microscopic states of a system [177]. Ensembles can be defined for any possible set of external constraints. The best known ones are the microcanonical (fixed number of particles N, volume V, and total energy E), canonical (fixed N, V and temperature T), the isoenthalpic-isobaric (fixed N, pressure P, and entropy S), the isothermal-isobaric (fixed N, P, T), often called NPT, and the grandcanonical (fixed V, T, and chemical potential μ) ensembles. Because most experiments are performed in the isothermal-isobaric ensemble we will focus on this one in the following.

As a start, I briefly summarize the laws of thermodynamics. The first law states that the internal energy change ΔU

$$\Delta U = Q + W \tag{2.1}$$

results from the work W the surrounding performs on the system and the heat Q that is added to the system during a process. This first law basically describes the energy conservation, i. e. energy is only transformed from one form into another but can never be destroyed or created.

The second law of thermodynamics discusses the total entropy S (a state variable), also called disorder, of a system. It states that S cannot decrease in isolated systems, i. e. when there is no exchange of energy or matter with the surrounding. Another wording is that the heat from a warmer body will naturally flow to a colder one. The second law reads

$$\Delta S \geq \frac{Q}{T} \tag{2.2}$$

where T is the temperature. This means that the change in entropy equals the amount of heat Q added to the system with respect to the temperature for reversible processes. For irreversible processes the entropy exceeds it.

2.2.1 The isothermal-isobaric ensemble and the Gibbs free energy

One can think of isobaric systems as being coupled to external pistons that enable to compress or expand the volume of the system in order to keep the internal pressure P constant. Analogously, isothermal systems can be viewed as being coupled to external thermal reservoirs exchanging heat with the system in order to keep the internal temperature T constant. The respective state function or thermodynamic potential of the isothermal-isobaric ensemble is the *Gibbs free energy* $G(N, P, T)$ which is connected to the particle number N, the pressure P, and the temperature T via

$$\mathrm{d}G = \mu \mathrm{d}N + V\mathrm{d}P - S\mathrm{d}T \tag{2.3}$$

where μ is the chemical potential, V is the volume and S is the entropy. G is labeled "free" energy because it denotes the amount of energy which is "free" to perform work at a constant temperature T and pressure P [178]. The term was coined by Hermann von Helmholtz [179]. The Gibbs free energy is related to the isothermal-isobaric partition function Δ by

$$G(N,P,T) = -\frac{1}{\beta} \ln \Delta(N,P,T) \tag{2.4}$$

where $\beta = 1/(k_\mathrm{B}T)$ is the inverse temperature with the Boltzmann constant k_B. For a mathematical derivation and further background on ensembles, their partition functions, and their interconnection using Legendre transformations the reader is referred to [176]. Knowing the Gibbs free energy G and its change ΔG_AB between two system states A and B directly enables one to calculate many macroscopic observables of a given isothermal-isobaric system and is thus key to understanding macroscopic systems and processes. These are, for example, the enthalpy $H = -\partial \ln(\Delta(N,P,T))/(\partial\beta)$, the heat capacity $C_P = \partial H/\partial T$, and the chemical potential $\mu = (\partial G/\partial N)_{P,T}$ to name just a few.

The change in free energy (the free energy difference) can tell if energy must be added to the system in order for a reaction to happen or if it occurs spontaneously. The free energy difference tells, for example, if a chemical compound is a promising drug candidate [176] or if two protein domains form a stable interaction. The standard binding free energy difference ΔG^0_bind is also directly connected to experimental observables such as the equilibrium dissociation constant K_D

$$\Delta G^0_\mathrm{bind} = -k_\mathrm{B}T \ln(c_0 K_\mathrm{D}) \tag{2.5}$$

with the standard concentration $c_0 = 1\,\mathrm{mol\,L^{-1}} \approx 1/1661\mathrm{\AA}^3$.

It is possible to express the free energy as a function of reaction/generalized coordinates of the system, often also called collective variables (CVs). Depending on the process one is interested in, these can be, for example, angles, distances, and root mean squared deviations (RMSDs), or also a parameter of the system Hamiltonian such as Lennard-Jones interactions. These coordinates span free energy hyper surfaces, which contain information about stable conformers of the system (with respect to the chosen coordinates) and their relative stability, barriers, and minimum free energy paths in-between them. The derivative of the free energy along a CV is the ensemble-averaged force [180], i.e. the force along the CV averaged over all configurations of a system, whereby the free energy along a CV is often called a *potential of mean force* (PMF). The average force is the first part of the instantaneous force acting along the CV, the second part is a random force with zero average. It includes the fluctuations of all additional degrees of freedom and thus enforces the progression of the CV. These dynamics take place along the time-independent PMF [181]. Generally one can obtain the free energy along a reaction coordinate ξ from the probability density function $P(\xi)$ of the CV using

$$G(\xi) = -k_\mathrm{B}T \ln P(\xi) \text{ [182]} \tag{2.6}$$

if the sampling of the phase-space is sufficient, i.e. the simulation is sufficiently long and was not trapped inside local free energy minima. However, this does not mean that the resulting free energy is a meaningful quantity. If ξ was chosen poorly then $G(\xi)$ will be meaningless in terms of describing the actual system states. There are two major requirements for CVs in molecular dynamics simulations. The relevant metastable states as well as the transition states between them have to be distinct regions in CV space. Thus, they have to be energetically separate regions in the chosen space. In contrast, if different metastable states are projected onto the same CV space, energy barriers are integrated out and major sampling problems occur when using CV-based sampling methods discussed below [183].

2.2.2 Jarzynski equality

In this section the canonical ensemble with the respective state function F, the Helmholtz free energy, is used for the derivation of the Jarzynski equality [184] since this is more convenient. The results can easily be extended to the isothermal-isobaric ensemble and the Gibbs free energy G. The Jarzynski equality allows to calculate equilibrium properties from non-equilibrium simulations.

To understand the issue when performing non-equilibrium simulations we start by multiplying T to both sides of eqn. (2.2), the second law of thermodynamics, and using eqn. (2.1), the first law of thermodynamics, to substitute Q and we obtain

$$T\Delta S \geq \Delta U - W \tag{2.7}$$

which is equal to

$$W \geq \Delta U - T\Delta S \,. \tag{2.8}$$

Since the right hand side of the equation is the Helmholtz free energy $\Delta F = \Delta U - T\Delta S$, the maximum work theorem

$$W \geq \Delta F \tag{2.9}$$

is obtained. The external work that is done on a system is

$$W(t) = \int_0^t \frac{\partial H(\mathbf{z}(t'), t')}{\partial t'} \mathrm{d}t' \text{ [182]} \tag{2.10}$$

where $\mathbf{z}(t) \equiv (\mathbf{q}(t), \mathbf{p}(t))$ is a point in phase phase. When a process is studied where some parameters of the system undergo a change in time, this process can be described using a parameter λ, a switching time τ and the respective Hamiltonian H_λ as

$$W(\tau) = \int_0^\tau \frac{\partial \lambda(t)}{\partial t} \frac{\partial H_\lambda}{\partial \lambda}(\mathbf{z}(t)) \mathrm{d}t \text{ [184].} \tag{2.11}$$

The maximum work theorem then reads

$$\langle W(\tau) \rangle \geq \Delta F \tag{2.12}$$

where $\langle \cdots \rangle$ is the ensemble average over all paths from the initial to the final state in time τ and the equality only holds, when the system undergoes a reversible process ($\tau \to \infty$) or is in equilibrium because then the entropy can be constant. Jarzynski then showed, that the inequality in eqn. (2.12) can be transformed into an equality

$$e^{-\beta \Delta F} = \langle e^{-\beta W(\tau)} \rangle \tag{2.13}$$

when considering the exponential of the work instead of the work itself [184]. Therefore the Jarzynski equality enables to perform non-equilibrium simulations and calculate equilibrium properties from them. That is useful for alchemical simulations.

It is possible to generalize the Jarzynski equality to the isothermal-isobaric ensemble [185, 186] which then reads

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta G} \,. \tag{2.14}$$

2.3 Molecular mechanics and force fields

Molecular and cell biological systems are quantum mechanical systems in nature. At the present day it is still not computationally feasible to study most processes involving biomolecules extensively using quantum mechanical methods. For this reason approximations have been derived. The most common one is the Born-Oppenheimer approximation [187] to the Schrödinger equation. It states that the wave function can be split into an electronic and a nuclear wave function which can be described separately, because the masses of the nuclei of atoms are much larger than the masses of all electrons in the atoms. Therefore, the electronic wave function only depends on the nucleic positions but not on their momenta. Molecular mechanics (MM) as a classical method is implicitly based on this approximation but goes even further by treating atoms as perfect balls which are connected into molecules via harmonic springs to describe covalent bonds between them as well as other force terms that keep the angles and dihedrals in a molecule nearby certain resting positions [188]. The electrons are not included explicitly but their effect is rather taken into account by allowing different electron shell configurations for the various atoms of the same chemical element. This is implicitly implemented by utilizing different parameters for the same atoms when they are in a certain chemical surrounding [189]. These parameters are defined in the so-called MM force fields (FFs) that represent collections of parameters for various atoms, bonds, angles and dihedrals between them and additionally parameters for non-bonded interactions such as Van der Waals or electrostatic interactions. The synthesis of all parameters for all atoms in a certain molecule is defined in the basic potential energy function

$$E = \sum_{\text{bonds}} E_{\text{stretch}} + \sum_{\text{angles}} E_{\text{bend}} + \sum_{\text{dihedrals}} E_{\text{torsion}} + \sum_{\text{pairs}} E_{\text{nonbond}} \,. \tag{2.15}$$

The individual energy terms are the contributions from bond stretching (E_{stretch}), the bending of angles (E_{bend}), the rotations or torsions around specific bonds (E_{torsion}) and the non-bonded interactions between atoms or groups of atoms which are not directly connected via bonded terms ($E_{\text{nonbond}} = E_{\text{VdW}} + E_{\text{electrostatic}}$) [188]. Some FFs like the common CHARMM FFs [190–192] include additional terms like an improper dihedral term ($E_{\text{impr.dihedral}}$) which forces atoms to remain in a plane or the Urey-Bradley term (E_{UB}) which keeps the not directly interacting atoms in the angle terms at a certain distance to each other. The specific functions for the energy terms in the

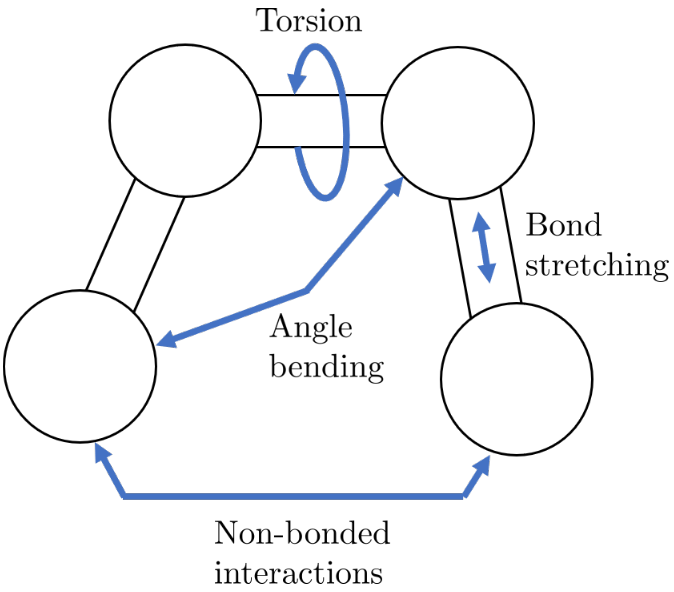


FIGURE 2.2: Schematic view of a molecule and the most common molecular mechanics terms.

CHARMM FFs have the form [190]

$$\begin{aligned}
E_{\text{stretch}} &= K_b(b-b_0)^2 \\
E_{\text{bend}} &= K_\theta(\theta-\theta_0)^2 \\
E_{\text{torsion}} &= K_\chi(1+\cos(n\chi)-\delta) \\
E_{\text{VdW}} &= \epsilon_{ij}\left[\left(\frac{R_{\text{min},ij}}{r_{ij}}\right)^{12}-\left(\frac{R_{\text{min},ij}}{r_{ij}}\right)^{6}\right] \\
E_{\text{electrostatic}} &= \frac{q_i q_j}{4\pi D r_{ij}} \\
E_{\text{impr.dihedral}} &= K_{\text{imp}}(\phi-\phi_0)^2 \\
E_{\text{UB}} &= K_{\text{UB}}(S-S_0)^2\,.
\end{aligned}$$

The constants K_X in every term are chosen based on the involved atoms in a bond, angle, dihedral etc. and usually fitted to resemble properties which have been previously calculated using quantum mechanical (ab initio or density functional) methods or experimentally measured. This is called parametrizing. Recent versions of the CHARMM FFs include additional terms to include grid-based energy corrections maps (CMAP) which are cross terms for the backbone dihedral angle values of protein backbones to allow for better reproduction of the complete quantum mechanical potential energy surfaces [191].

The main advantage of MM is, that it requires much less computer power to simulate molecular systems compared to quantum mechanical methods. Therefore, it is thus suitable to study large molecules like proteins by molecular dynamics simulations involving millions to billions of force calculations for the entire system. This efficiency comes at the price of the accuracy of the results, which are highly dependent on the force field parameters, and the impossibility of determining electronic properties and describing bond breaking or forming.

2.4 Molecular dynamics simulations

There are two main methods when it comes to simulating molecular and biological systems. These are the Monte-Carlo (MC) and molecular dynamics (MD) methods. In the former one the Metropolis algorithm [193] is used in that the positions of the atoms in the system are propagated randomly in a fixed search radius in every simulation step. Afterwards the energy difference ΔE between the old and new state is calculated and the new configuration is accepted with a probability $p = \min(1, \exp(-\Delta E/(k_B T)))$ where T is the temperature of the system and k_B is the Boltzmann constant. Due to the random sampling the MC method is not suitable to study the dynamics of systems. Which is why the MD method is used most commonly to study molecular systems. Here the classical Newton's equations of motion for many-body systems are solved numerically with respect to boundary conditions that are appropriate for the system of choice and the parameters describing the particle-particle interactions which are stored in the previously described force fields (FFs). Thus, MD simulations enable to study the microscopical time evolution of molecular and biomolecular systems and can therefore give insights into the dynamics of these systems and the respective properties such as rate constants or transport properties. At the same time they also allow to obtain equilibrium properties like free energies along reaction coordinates, e. g. binding free energies, and macroscopic thermodynamic properties such as pressure, temperature and volume by sampling from the statistical mechanical ensemble [194]. Since classical FFs cannot describe bond breaking or forming, MD simulations can be combined with electronic structure calculations like density functional theory [195] calculations to account for the quantum mechanical nature of the electrons involved in atomic interactions. Methods such as path integral molecular dynamics [196] enable to account for nuclear quantum effects as well which can be important in processes like proton transfer reactions where the classical approximation for the nuclei is not valid anymore. These extensions remove the dependency of MD simulations on approximated FF parameters. However, they introduce a large computational overhead and are therefore not suited to nowadays being generally applied in simulations of biomolecular systems on common simulation hardware.

The basis of MD simulations are Hamiltonian mechanics where

$$H(\mathbf{q},\mathbf{p},t) \equiv H(\mathbf{p}_1(t),\cdots,\mathbf{p}_N(t),\mathbf{q}_1(t),\cdots,\mathbf{q}_N(t),t) = \sum_{i=1}^{N} \frac{(\mathbf{p}_i(t))^2}{2m_i} + U(\mathbf{q}_1(t),\cdots,\mathbf{q}_N(t)) \tag{2.16}$$

is the Hamiltonian $H(\mathbf{q},\mathbf{p},t) = H(\mathbf{q}(t),\mathbf{p}(t),t)$ for an N-particle system where only inter-particle interactions are considered. The explicit mentioning of time dependence is omitted in the following for the ease of reading. $\mathbf{q}_1,\cdots,\mathbf{q}_N$ are the spatial particle positions at time t and $\mathbf{p}_1,\cdots,\mathbf{p}_N$ with $\mathbf{p}_i = m_i\mathbf{v}_i$ are the momenta and $\mathbf{v}_1,\cdots,\mathbf{v}_N$ are the velocities of these particles while $U(\mathbf{q}_1,\cdots,\mathbf{q}_N)$ is the inter-particle potential. The positions and momenta can be combined into a so-called phase space vector $\mathbf{x} = (bp_1,\cdots,\mathbf{p}_N,\mathbf{q}_1,\cdots,\mathbf{q}_N)$ and all possible phase space vectors combined are forming the $2dN$-dimensional phase space (d is the number of spatial dimensions) where every phase space vector describes a classical state of the system [194]. The resulting Newton's equations of motion from Newton's second law

$$m\ddot{\mathbf{q}}_i = \mathbf{F}_i \tag{2.17}$$

with the forces $\mathbf{F}_1, \cdots, \mathbf{F}_N$ on the N particles at time t can be obtained by application of Hamilton's equations

$$\dot{\mathbf{q}}_i = \frac{\partial H(\mathbf{q}, \mathbf{p})}{\partial \mathbf{p}_i} = \frac{\mathbf{p}_i}{m_i} \tag{2.18}$$

$$\dot{\mathbf{p}}_i = -\frac{\partial H(\mathbf{q}, \mathbf{p})}{\partial \mathbf{q}_i} = -\frac{\partial U}{\partial \mathbf{q}_i} = \mathbf{F}_i(\mathbf{q}_1, \cdots, \mathbf{q}_N) \,. \tag{2.19}$$

The Hamiltonian $H(\mathbf{q}, \mathbf{p})$ is conserved by these equations of motion and these therefore generate microscopic configurations which belong to a microcanonical ensemble with energy $H(\mathbf{q}, \mathbf{p}) = \text{const.} = E$ [176]. Additionally, these equations of motions are time reversible.

If all possible configurations described by the Hamiltonian can be visited in an infinite amount of time, the system is said to be ergodic. In this case time averaging can replace the microcanonical phase-space average for any observable a

$$\langle a(\mathbf{q}, \mathbf{p}) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_0^T \mathrm{d}t \; a(\mathbf{q}, \mathbf{p}, t) \text{ [176].} \tag{2.20}$$

Eqn. (2.18) and eqn. (2.19) are solved numerically in MD simulations with respect to a number of initial conditions which is the reason why a numerical integration scheme is needed. The most commonly known algorithms are the Verlet [197], velocity Verlet [198] and leap-frog [199] algorithms, which will be discussed in the following.

2.4.1 Numerical integration schemes - Finite-difference methods

In order to numerically integrate a continuous equation it has to be transformed into a discrete form. In order to do this for the equations of motion eqn. (2.18) and eqn. (2.19), a discrete time step Δt has to be introduced which allows to propagate the equations in time. The time step has to be chosen small enough to allow the sampling of all relevant degrees of freedom in the studied systems but large enough that it is still computationally feasible to sample a certain amount of time. Thus, it is always a trade off between accuracy and computational effort whereby a common choice for MD simulations is $\Delta t = 2\,\text{fs}$.

Numerical integration methods are often derived by a Taylor expansion of the relevant degrees of freedom at time $t + \Delta t$. Accordingly, for the position $\mathbf{q}_i$ of particle i at time $t + \Delta t$ the Taylor expansion with terms higher than second order dropped reads

$$\mathbf{q}_i(t + \Delta t) = \mathbf{q}_i(t) + \Delta t \; \dot{\mathbf{q}}_i(t) + \frac{1}{2}\Delta t^2 \; \ddot{\mathbf{q}}_i(t) + O(\Delta t^3) \tag{2.21}$$

or

$$\mathbf{q}_i(t + \Delta t) \approx \mathbf{q}_i(t) + \Delta t \; \mathbf{v}_i(t) + \frac{\Delta t^2}{2m_i}\mathbf{F}_i(t) \tag{2.22}$$

because $\dot{\mathbf{q}}_i(t) = \mathbf{v}_i(t)$ is the velocity of the particle at times t and $\ddot{\mathbf{q}}_i(t) = \frac{\mathbf{F}_i(t)}{m_i}$ is the force acting on the particle at time t. Here, we dropped all higher order error terms. The same can be done by considering the Taylor expansion for the position $\mathbf{q}_i(t - \Delta t)$

$$\mathbf{q}_i(t - \Delta t) \approx \mathbf{q}_i(t) - \Delta t \; \mathbf{v}_i(t) + \frac{\Delta t^2}{2m_i}\mathbf{F}_i(t) \tag{2.23}$$

and if both equations (eqn. (2.22) and eqn. (2.23)) are added, the velocity-independent scheme

$$\mathbf{q}_i(t+\Delta t) \approx 2\mathbf{q}_i(t) - \mathbf{q}_i(t-\Delta t) + \frac{\Delta t^2}{2m_i}\mathbf{F}_i(t) \tag{2.24}$$

is obtained. This is commonly known as the Verlet algorithm [197] and depends only on the positions at time t and $t - \Delta t$ and the force acting on the particle. The respective velocity of the particle

$$\mathbf{v}_i(t) = \dot{\mathbf{q}}_i(t) \approx \frac{\mathbf{q}_i(t+\Delta t) - \mathbf{q}_i(t-\Delta t)}{2\Delta t} + O(\Delta t^2) \tag{2.25}$$

is obtained by subtraction of eqn. (2.23)) from eqn. (2.22) and can therefore also be calculated even though they are not directly evolved in time but only calculated from the change in the positions [176]. This is called the centered difference for the first derivative that has an accuracy $O(\Delta t^2)$. It is also possible to obtain a forward and backward difference for the first derivative

$$\dot{\mathbf{q}}_i(t) = \frac{\mathbf{q}_i(t+\Delta t) - \mathbf{q}_i(t)}{\Delta x} + O(\Delta t) \tag{2.26}$$

that can be derived from eqn. (2.22) and eqn. (2.23), respectively, and has an accuracy of $O(\Delta t)$ wherefore it is more common to use the centered difference.

To also propagate the velocity explicitly at every time step, another scheme, the so-called velocity Verlet algorithm [198], can be derived. Here eqn. (2.23) is expanded around t instead of $t - \Delta t$

$$\mathbf{q}_i(t) \approx \mathbf{q}_i(t+\Delta t) - \Delta t\,\mathbf{v}_i(t+\Delta t) + \frac{\Delta t^2}{2m_i}\mathbf{F}_i(t+\Delta t) \tag{2.27}$$

into which $\mathbf{q}_i(t+\Delta t)$ from eqn. (2.22) can be inserted which results in

$$\mathbf{v}_i(t+\Delta t) \approx \mathbf{v}_i(t) + \frac{\Delta t}{2m_i}\left(\mathbf{F}_i(t) + \mathbf{F}_i(t+\Delta t)\right) \tag{2.28}$$

where the velocity is propagated from time t to $t + \Delta t$ using the velocity at time t and the forces acting on the particle at time t and $t + \Delta t$. This can be used in combination with eqn. (2.22) to propagate the position $\mathbf{q}$ and the velocity $\mathbf{v}$ at the same time. The usual implementation is

1. Calculate the velocity $\mathbf{v}_i(t+1/2\Delta t) = \mathbf{v}_i(t) + \frac{\Delta t}{2m_i}\mathbf{F}_i(t)$
2. Calculate $\mathbf{q}_i(t+\Delta t) = \mathbf{q}_i(t) + \mathbf{v}_i(t+1/2\Delta t)\Delta t$
3. Calculate the new force $\mathbf{F}_i(t+\Delta t)$ using the new position $\mathbf{q}_i(t+\Delta t)$
4. Calculate the final velocity $\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t+1/2\Delta t) + \frac{\Delta t}{2m_i}\mathbf{F}_i(t+\Delta t)$.

If the force does not depend on the velocity of the particle then the first step can be ignored and the velocity can be directly calculated as in eqn. (2.28).

Instead of integrating the position and velocity at the same time, like in the velocity Verlet algorithm [198], it is also possible to propagate the position at time t, $t + \Delta t$, $t + 2\Delta t$, $\cdots$ and the velocity at time $t + 1/2\Delta t$, $t + 3/2\Delta t$, $t + 5/2\Delta t$, $\cdots$.

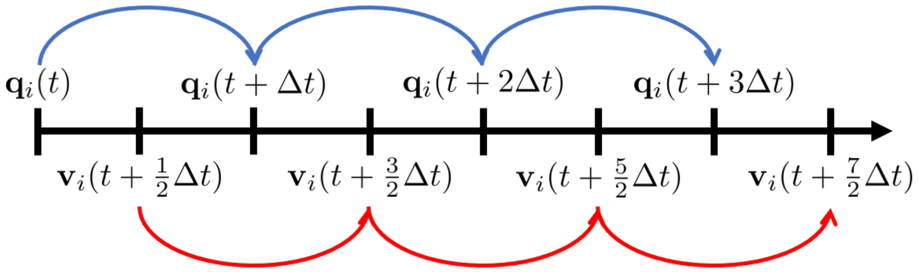


FIGURE 2.3: The leap-frog scheme illustrated.

Accordingly, the numerical equations of motion are

$$\mathbf{v}_i(t+1/2\Delta t) \approx \mathbf{v}_i(t-1/2\Delta t) + \frac{\Delta t}{2m_i}\mathbf{F}_i(t) \tag{2.29}$$

$$\mathbf{q}_i(t+\Delta t) \approx \mathbf{q}_i(t) + \Delta t\, \mathbf{v}_i(t+1/2\Delta t) \tag{2.30}$$

This scheme is called the leap-frog algorithm [199] and was used throughout this thesis. The name evolved due to the leapfrog-look of the scheme as can be seen in fig. 2.3. The velocities at integer multiples of the time step Δt can again be calculated using

$$\mathbf{v}_i(t) = \frac{\mathbf{v}_i(t+\Delta t) + \mathbf{v}_i(t-\Delta t)}{2} . \tag{2.31}$$

All of these algorithms are time reversible, which is a fundamental property of the Hamiltonian equations of motion (eqn. (2.18) and eqn. (2.18)) and is therefore crucial to be fulfilled by a numerical integrator for these equations. Time reversibility means that if the system is propagated n steps forward and then n steps backward in time it will end up in the same state it started from. The second important property of the Hamiltonian equations of motions which is fulfilled by these algorithms is symplecticity, i. e. the preservation of volume in phase space.

It is also possible to use higher-order integration methods for higher accuracy of the integrator, though these methods come at the price of increased computational costs. Higher accuracy can also be achieved by reducing the time step Δt which is usually less computational expensive compared to using a higher-order scheme.

2.4.2 Periodic boundary conditions

Due to computational limitations [200] only a small part of the real system, a unit cell, is simulated in usual MD simulations. Even very large MD simulations with millions of atoms are not even close to the amount of atoms in a realistic system (e. g. 10 million atoms are only 1.66×10^{-17} mol). The extension to the real (on molecular scales infinite) system is reached using the so-called periodic boundary conditions (PBCs). Their application can be understood as an infinite repeat of images of the unit cell in all directions which is equivalent to an infinite Bravais lattice where the shape of the simulation box determines the type of the lattice [201]. Here, one assumes that the simulated system is periodic in all directions [193] which is reasonable since diffusion usually leads to an equal distribution of substances. When a particle leaves the unit cell on one side it moves into a copy of the same unit cell and thus it is introduced to the system again on the opposite side of the box. In order to keep the correct interactions between the particles in the system, each particle interacts with the closest copy (or closest image) of an interaction partner no matter if it is in the real cell or in a copy of the cell. This is the so-called minimum-image convention [193]. However, the periodicity can still lead to artifacts if molecules interact with their periodic images e. g. due to long-range electrostatic interactions. These artifacts can be reduced by increasing of the simulation box size [202].

2.4.3 Electrostatic interactions and lattice-sum methods

Properly treattin electrostatic interactions, one type of the so-called non-bonded interactions, in MD simulations is a challenging task because they are usually long-ranged. On the one hand, being proportional to the inverse distance r^{-1} between the charges they have to be approximated because their exact application would lead to an enormous computational effort. On the other hand, every approximation has a strong effect on the accuracy of the simulation [203]. The total Coulomb interaction energy reads

$$V^{\mathrm{Coul}} = \frac{1}{4\pi\epsilon_0} \sum_{i,j} \frac{q_i q_j}{|\mathbf{r}_{ij}|} \tag{2.32}$$

where $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ is the distance between the positions $\mathbf{r}_i$ and $\mathbf{r}_j$ of ions i and j with charges q_i and q_j, respectively. ϵ_0 is the vacuum permittivity and the sum is evaluated over all pairs of ions. The easiest method to calculate electrostatic interactions is to define a fixed cutoff distance, typically in the range of $10 - 12$ Å beyond which the interactions are assumed to be zero [204]. This can lead to strong truncation errors when the simulazion system contains strong interactions. For simulation systems subject to PBCs the so-called lattice-sum methods [8, 9, 204–214] are the methods of choice since they assume the system to be a crystal consisting of an infinite lattice of periodic copies of itself [203] as it is created when using PBCs. This allows to calculate the long-range Coulomb interactions using techniques originally developed for crystals without being dependent on an unphysical change of the potential [204]. One of the most commonly used lattice-sum methods is the particle-mesh Ewald method [8] (PME). It was derived from the Ewald method [205] where the electrostatic energy term is split into a short-range and a long-range term. The former decays fast with increasing distance $|\mathbf{r}_{ij}|$ and can thus be cut off at a reasonable value of $|\mathbf{r}_{ij}|$. The latter is very smooth and expanded and therefore it is very concentrated in reciprocal space (or k-space, the Fourier transform of the real space) and can therefore be truncated at reasonable values of the wave vector k. In the Ewald method the lattice summations are performed analytically [204] and it has a runtime of N^2, with N being the number of atoms, due to the calculation of the reciprocal sum. In contrast, the particle-mesh Ewald method assigns charges to a grid using interpolation, since charges are usually not sitting directly on a grid point. The grid is then transformed to a reciprocal lattice using a Fast-Fourier transformation and the reciprocal sum is calculated as a single sum over the grid on the reciprocal lattice [8]. The usual implementation of the PME method uses B-splines instead of Lagrange interpolation which significantly improves the accuracy of the method and is usually called smooth PME [9]. Overall, that leads to a runtime of $N \log N$ that is therefore greatly reduced compared to the Ewald method.

2.4.4 Van-der-Waals interactions

The second type of non-bonded interactions are the so-called Van-der-Waals interactions which include repulsive, resulting from the Pauli exclusion principle, and attractive interactions. The latter can be divided into London dispersion interactions [215, 216], which are interactions between induced dipoles, dipole-dipole interactions and dipole-induced-dipole interactions. All these interactions are usually treated in a combined manner using the Lennard-Jones potential [217–219] in MD simulations. This consists of a repulsive term ($(\sigma/|\mathbf{r}_{ij}|)^{12}$) and an attractive term ($(\sigma/|\mathbf{r}_{ij}|)^{6}$) and reads

$$V_{ij}^{\mathrm{LJ}}(\sigma/|\mathbf{r}_{ij}|) = 4\epsilon_{ij}\left[(\sigma_{ij}/|\mathbf{r}_{ij}|)^{12} - (\sigma_{ij}/|\mathbf{r}_{ij}|)^{6}\right] \tag{2.33}$$

where $|\mathbf{r}_{ij}|$ is the distance between particles i and j, σ_{ij} represents the distance where the potential is zero for these particles and ϵ_{ij} is the depth of the potential for these particles and therefore the strength of the interaction.

Van-der-Waals interactions are usually treated by using a cut-off radius in MD simulations. In order to obtain continuous forces at the cut off force-switch functions are often used which switch the force to zero at the cut-off. The introduced error is usually much smaller than other errors in a simulation.

2.4.5 Solvation

Biomolecular systems nearly always involve water and all biological molecules are usually floating in solvents in our cells. Additionally, they are not only solvated but water molecules even play an important role for many reactions. For some purposes, it may be adequate to simulate molecules in gas phase, for example if one is interested in the molecule itself without interactions with the surrounding [188]. However, in order to understand biomolecular properties solvation can typically not be ignored.

Solvation can be applied computationally in two ways, that is in an implicit and an explicit way. In the latter case, solvent molecules are placed around the solute into the simulation box, in the former the solute is surrounded by a continuous medium [188] which resembles the properties of the solvent [220]. Explicit solvation is needed when it is likely that solvent molecules are directly involved in the study system. Implicit solvation, in contrast, mostly influences the charge distribution but does not allow the solute to directly interact with the solvent, though it has the advantage of drastically reduced computational effort since much fewer atoms have to be sampled during the simulation. We will focus on explicit solvation and especially on water models because we used explicit water as the only solvent throughout this thesis.

In order to accurately describe real biomolecular and chemical systems, an explicit water model has to accurately describe solute-solvent interactions, since these can strongly influence the properties of the solute. As important are solvent-solvent (or water-water) interactions, since solvent molecules usually account for over 80% of the simulation particles. Overall the water model should resemble experimental observations, namely thermodynamic and kinetic properties, of the water itself and for solutions. The model also has to match the method of treating the long-range electrostatic interactions, e. g. particle-mesh Ewald (see 2.4.3), since the treatment of the Coulomb interactions can have an effect on the macroscopic properties of the solute calculated using the simulations [221, 222].

Water models for MD simulations can be divided into three-, four- or higher-site models. Common 3-site water models are the Simple Point Charge (SPC) [10] and the SPC/E [223] (improved SPC by adding an polarization correction) water models and the TIP3P water model [224] with its special CHARMM FF derivative [190].

These models have three points of interaction with a point charge at each which resemble the three water atoms. Three-site models are the most commonly used ones since they are the most computationally efficient explicit water models.

Higher site models add dummy atoms near the oxygen of the water molecule which have negative charges and should account for the lone electron pairs on the oxygen. These should improve the description of the electrostatic interactions and the interactions of water molecules with each other. Commonly known higher-site water models are the four-site TIP4P [224] and OPC [225] water models and the five-site TIP5P [226] water model. The disadvantage is, the more sites a water model has, the more particles have to be sampled in the simulation and therefore the computational effort drastically increases with increasing number of sites.

In order to use fixed-geometry water models SPC or TIP3P and to have rigid water molecules, constraints have to be applied to the water molecules in the MD simulations. In the original SHAKE algorithm [227] a set of Lagrange multipliers has to be solved in order for the equations of motion to fulfill the constraints. The RATTLE algorithm [228] is an advancement to SHAKE since it also adjusts the velocities and not only the coordinates of the particles and is therefore suited to be used with the velocity Verlet or leap-frog algorithms (see section 2.4.1). These algorithms can be used to constrain bonds, angles and distances off pairs of atoms in any sense. Specifically for rigid water molecules the SETTLE algorithm [229] was created which is an analytical version of SHAKE and RATTLE whereby it is faster and has a higher accuracy than the other algorithms [229]. Therefore it was the method of choice for this thesis to constrain water molecules in the MD simulations. As stated above, solvent molecules make up around 80% of all particles in a simulation system and it is therefore essential to have fast methods specifically designed for these. Additionally, constraining of the water molecule bonds and angles allows to use a time step of 2 fs in the MD simulation because the constraints reduce the need to sample high frequency bond stretches and angle bends. Without the constraints usually only time steps of 1 fs or lower could be used that drastically increase the computational effort.

It is important to choose a water model suited for the task at hand because the choice of the water model can have a strong impact on the properties one tries to obtain from MD simulations. For example it was shown that the choice of the water model influences the stability of the Trp-cage miniprotein [230] as well as can suppress peptide folding by solvating the unfolded state [231]. Additionally, the water model can strongly influence the binding free energy of conserved water molecules in binding pockets [232] and also the protein-ligand electrostatic binding free energies [233].

2.4.6 Thermostatting and barostatting

If MD simulations are performed in a simulation box with fixed size (i. e.volume) and particle number using a suitable integrator and time step Δt to keep the total energy from drifting (i. e. keeping the total energy constant) the simulations sample the microcanonical ensemble. However, experiments are usually performed in the isothermal-isobaric ensemble [234] (see section 2.2.1) since this most often best resembles the real system. We will focus on this ensemble since it was used throughout this thesis. To perform simulations in the isothermal-isobaric ensemble thermostats and barostats have to be used.

Thermostat algorithms are used to keep the temperature constant in a simulation. These couple the system to an external heat bath with a reference temperature T_0 [234]. An instantaneous temperature $\mathcal{T}$ for the simulated system needs to be defined to be compared to the reference temperature during the simulation. One possible definition is

$$\mathcal{T} = \frac{2}{k_{\mathrm{B}} N_{\mathrm{df}}} \mathcal{K} \tag{2.34}$$

with the Boltzmann constant k_{B}, the number of internal degrees of freedom of the system N_{df} and the instantaneous internal kinetic energy $\langle \mathcal{K} \rangle$ [234]. This ensures that $\langle \mathcal{T} \rangle = T$, i. e. the average temperature in the simulation, $\langle \mathcal{T} \rangle$, is equal to the macroscopic temperature, T, which we want the system to be at, because the average kinetic energy of the system, K, fulfills

$$K = \langle \mathcal{K} \rangle = \frac{1}{2} k_{\mathrm{B}} N_{\mathrm{df}} T \text{ [234].} \tag{2.35}$$

For a deeper mathematical derivation of the internal degrees of freedom and the respective calculation of the kinetic energy and temperature the reader is referred to Hünenberger 2005 [234].

Various thermostatting schemes exist, e. g. the Langevin thermostat, which applies a frictional and a stochastic force to the particles. Other examples are the Andersen thermostat [235], where the velocities of system particles are recurringly drawn from a Maxwell-Boltzmann distribution with the desired temperature T, the Nosé-Hoover thermostat [236, 237], where the system is coupled to a friction from an (or multiple) external heat bath(s) with its own momentum, the Berendsen thermostat [238] and the velocity-rescaling thermostat by Bussi et al. [239] that was used throughout this thesis. The latter two will be explained in more depth since they follow a similar approach.

Rescaling thermostats operate by scaling the velocity of each system particle i with a factor λ that depends on the ratio of the instantaneous temperature $\mathcal{T}$ and the macroscopic temperature T which should be maintained during the simulation, e. g.

$$v_i^{\mathrm{new}} = \lambda v_i \, . \tag{2.36}$$

In the Berendsen thermostatting scheme the scaling factor is

$$\lambda = \left[1 + \frac{n_{\mathrm{TC}} \Delta t}{\tau_T} \left(\frac{T}{\mathcal{T}} - 1 \right) \right]^{\frac{1}{2}} \tag{2.37}$$

where n_{TC} is the number of time steps after which the thermostat is applied and t_T is the strength of the coupling that has to be carefully chosen. Since the Berendsen thermostat is not associated with a well-defined ensemble (i. e. it does not create a canonical ensemble), Bussi et al. derived an extension to this thermostat [239], also called the Bussi-Donadio-Parrinello thermostat or stochastic velocity rescaling. In this method, a stochastic term is added to ensure that the distribution for the kinetic energy is correct. The correction for the instantaneous kinetic energy is

$$\mathrm{d}\mathcal{K} = [K - \mathcal{K}(t)] \frac{\mathrm{d}t}{\tau_T} + 2 \sqrt{\frac{\mathcal{K}(t) K}{N_{\mathrm{df}} \tau_T}} \xi(t) \tag{2.38}$$

or the respective correction for the temperature

$$\mathrm{d}\mathcal{T} = [T - \mathcal{T}(t)] \frac{\mathrm{d}t}{\tau_T} + 2 \sqrt{\frac{\mathcal{T}(t) T}{N_{\mathrm{df}} \tau_T}} \xi(t) \tag{2.39}$$

where ξ is a white noise term, the derivative of a Wiener process. The velocity is then rescaled with a factor

$$\alpha = \left[e^{\frac{-\Delta t}{\tau_T}} + \frac{K}{N_{\mathrm{df}}\mathcal{K}} \left(1 - e^{\frac{-\Delta t}{\tau_T}}\right) \left(R_1^2 + \sum_{i=2}^{N_{\mathrm{df}}} R_i^2\right) + 2e^{\frac{-\Delta t}{2\tau_T}} \sqrt{\frac{K}{N_{\mathrm{df}}\mathcal{K}} \left(1 - e^{\frac{-\Delta t}{\tau_T}}\right) R_1} \right]^{\frac{1}{2}} \quad [239] \tag{2.40}$$

as

$$v_i^{\mathrm{new}} = \alpha v_i \, . \tag{2.41}$$

It was additionally shown that this thermostat can be considered as a global version of the Langevin thermostat [240]. The stochastic velocity rescaling has another strong advantage when compared to the Berendsen thermostat or the original velocity rescaling. The use of the latter two can lead to the so-called "flying ice cube" effect, first described by Harvey et al. [241]. This effect is understood to mean that the system freezes with time, meaning that the energy of high-frequency modes is transferred to low-frequency modes and therefore the system gains increasing center-of-mass motions and looses energy on internal degrees of freedom. The researchers showed that this violates the equipartition principle [241, 242].

Another issue with thermostatting in MD simulations is the so-called "hot solvent - sold solute problem". Here the algorithmic noise of the thermostat can affect a strongly polar solvent more strongly than a mildly polar solute when using an electrostatic cut-off [242–246] This can occur when there are different sets of degrees of freedom with different frequencies or heating rates, that differ strongly due to algorithmic noise. A solution is to couple the solute and solvent each to a separate thermostat [234].

When only applying thermostat algorithms MD simulations will sample the canonical ensemble (even though not all thermostat algorithms can fulfill this, as stated before). When the simulations should sample the isothermal-isobaric ensemble then barostats have to be used in addition to the thermostat algorithms. The Berendsen barostat [238] is very similar to the previously described Berendsen thermostat. But instead of coupling the system to an external heat bath, it is coupled to an external pressure bath and not the velocities but the simulation box vectors and the coordinates are scaled periodically [247]. But again the sampled ensemble is not well defined and therefore the Berendsen algorithm should not be used for production runs in an MD simulation. Instead, it can be very successfully be used for equilibration because it can bring the system to the wanted pressure very fastly. Furthermore, there exist multiple other methods like the Andersen barostat [235], the Martyna-Tuckerman-Tobias-Klein barostat [248, 249] or the Parrinello-Rahman barostat [250], which is an extension of the Andersen barostat and also allows shape changes. The last was used throughout this thesis for the production runs of the simulations.

2.5 Overview over computational methods to determine binding free energies

In this section, an overview over computational methods that allow to determine binding free energies is given. Several concepts are shown and the concepts which are used in this thesis are presented in detail.

In principle it is possible to obtain binding free energies using unbiased long molecular dynamics (MD) simulations. Most commonly, however, the time scales of binding and unbinding are too long to be sampled on common simulation hardware.

In recent years specialized hardware was developed which allowed to conduct milliseconds long trajectories [251, 252] of solvated protein systems in atomistic detail. Notably, such simulations were recently applied to study the reversible association and dissociation kinetics of the five protein-protein complexes barnase-barstar, insulin dimer, ras-raf RBD, RNase HI-SSB-Ct, and TYK2-pseudokinase by atomistic MD simulations in explicit solvent [253]. The authors performed dozens of conventional MD simulations with aggregated simulation times of hundreds of microseconds. In addition, they performed "tempered" binding simulations, whereby "the strength of interactions between the protein monomer atoms, and sometimes between the protein monomer and solvent atoms, [was] scaled at regular time intervals using a simulated Hamiltonian tempering framework" [253]. This scaling was adjusted to allow dissociation from long-lived bound states to occur within hundreds of microseconds rather than days. Observing reversible binding and unbinding events yielded the following general picture. One may have thought that associating proteins could in principle form an encounter complex at an arbitrary interface and proceed from there to the native interface without dissociating by means of an extensive search [253]. Instead, the authors observed that "in successful association events the encounter complexes tended to form rather close to the native interface". On the other hand, encounter complexes that would not later reach the native interface formed in a wide variety of relative orientations. At present this specialized hardware is too expensive for the normal scientist to work with and other methods have been developed in order to deal with the sampling problem. In the following we will give an overview of different simulation methods, so-called enhanced sampling methods, which allow the system to overcome barriers in free energy, thus moving out of local or the global minima, which then enables one to calculate binding free energies *in silico*.

There exist two major forms of binding free energy methods. Methods that are used to calculate absolute binding free energies, i. e. the change in free energy that results when a ligand binds to a receptor, and methods used to calculate relative binding free energies, i. e. the difference in absolute binding free energy between two ligands binding to the same receptor. Even though both are usually called free energies, the former is a free energy difference between the bound and unbound states and the latter is even a difference of free energy differences. In the following common methods will be introduced shortly while the focus is placed on the methods used for this thesis.

2.5.1 Relative binding free energies using alchemical free energy simulations

Alchemical free energy methods are a special form of free energy perturbation approaches termed alchemical since they involve unnatural changes of the system structure. This can be atom creation, annihilation and morphing. Such methods usually involve a so-called thermodynamic cycle (fig. 2.4) defining a path between the initial state A and final state B of the system.

There are two types of alchemical free energy approaches, namely absolute and relative methods whereby the latter ones will be called alchemical transformations in the following. Absolute free energy simulations involve restraining the ligand in the binding pocket, releasing the interactions between the ligand and the receptor as well as turning on the interactions of the ligand in the free state (only surrounded by solvent). Additionally the restraints in the free state have to be removed. These

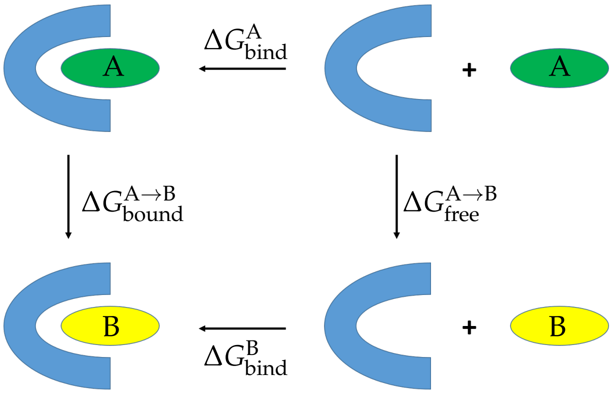


FIGURE 2.4: Thermodynamic cycle for the difference in binding free energy between ligands A and B. The figure was taken from ref. [32]. These two vertical processes are alchemical transformations.

types of simulations enable to calculate the absolute binding free energy between the ligand and the receptor.

In contrast, relative alchemical free energy simulations (or alchemical transformations) involve mutations of parts of the system. Two states A and B can be defined whose structures vary at these specific positions which should be mutated. This could be two ligands of a receptor differing in some parts (e. g. a peptide with alanine at a specific position in state A which should be transformed to a serine in state B). These transformations have to be performed in the bound and in the free state. In fig. 2.4 the thermodynamic cycle of an alchemical transformation with two different ligands A and B is drawn. It shows the four states of the simulation (bound states with ligand A and B, free states with ligands A and B) and the differences in free energy between them. This directly allows to calculate the difference in binding free energy between the two states A and B from the mutation free energies $\Delta G^{\mathrm{A}\to\mathrm{B}}_{\mathrm{bound}}$ and $\Delta G^{\mathrm{A}\to\mathrm{B}}_{\mathrm{free}}$ using the equation

$$\Delta\Delta G^{\mathrm{B-A}}_{\mathrm{bind}} = \Delta G^{\mathrm{B}}_{\mathrm{bind}} - \Delta G^{\mathrm{A}}_{\mathrm{bind}} = \Delta G^{\mathrm{A}\to\mathrm{B}}_{\mathrm{bound}} - \Delta G^{\mathrm{A}\to\mathrm{B}}_{\mathrm{free}}\,. \tag{2.42}$$

From now on the alchemical transformations will further be discussed using the example of two distinct peptides binding to a protein, since this is the particular application used during this thesis.

Calculating absolute free energies for the binding of peptides to a protein using pathway methods suffers from convergence problems because peptides are highly flexible amino acid chains and thus they involve many degrees of freedom which all have to be properly sampled in order to reach converged results. The alchemical transformation method does not suffer from these types of issues since the peptide will not be pulled away from the receptor and the interactions between the two partners will not be completely turned on or off. Only a (smaller) change in the peptide will be simulated. The topology is transformed from one state to the other one (A → B, B → A) by the use of a coupling parameter λ that is changing from 0 to 1. The resulting binding free energy difference is then calculated using eqn. (2.42). It is possible to perform these changes fast, i e. pulling the system away from equilibrium, or slow, i e. trying to stay in equilibrium. In the former case, only the initial state A has to be in (thermal) equilibrium and the rest of the simulation does not. Then the correct free energy is obtained by applying the Jarzynski equality [184] (see sec. 2.2.2) or the Crooks fluctuation theorem [254]. In these methods, ensemble averages have to be calculated and therefore a number of simulations have to be performed to obtain enough sampling. In contrast, for the slowly changing or equilibrium approach the change in λ has to be performed continuously very slowly or in fixed windows so that the system can always adapt to these changes. A common method for analysing the outcome is the so-called thermodynamic integration [255] (see sec. 2.5.3) that can also be used as a pathway method when λ is used to change a geometrical coordinate.

Even though relative free energy alchemical simulations usually do not suffer from the same convergence issues as absolute free energy methods they are not free of obstacles. Problems can arise when the two states A and B have a different total charge. In that case the overall system will not be charge neutral during the course of the simulation. That introduces errors due to the finite size of the simulation system, which is a common approximation when simulating biomolecular systems. As already described in section 2.4.2, the infinite non-periodic realistic system cannot be computationally handled in an explicit-solvent molecular dynamics simulation. Thus, finite simulation boxes with periodic boundary conditions (PBCs, see 2.4.2) are used to represent the real system. Electrostatic interactions in these systems are simulated using lattice-sum methods [8, 9, 200, 204–214] such as the particle-mesh Ewald method [8, 9] which are described in more detail in section 2.4.3. Artifacts can arise which are related to the Wigner self-energy [208, 213, 256–259] of a charge in a periodic system. Here a charge is interacting with its own copies in the periodic images and a homogeneous neutralizing background charge density [213]. A second type of artifact arises because the average potential in a system under PBC is zero while in a real setup the potential would only be zero at infinity [214, 260–266].

There are different possibilities to perform alchemical transformations. A common way is to simulate the bound and free legs of the thermodynamic cycle in separate boxes. The second possibility is to simulate both legs in a single box where the bound and free states are separated by a sufficiently large distance filled with solvent in order to keep them from interacting with each other. This is the so-called "double system in a single box setup" [267]. The different setups will be explained in the next sections.

Double system in a single box setup

In the "double system in a single box setup" [267] the bound and free states of the system are simulated together. For one of them the mutation is performed in the forward (A $\rightarrow$ B) and the other one in the backward direction (B $\rightarrow$ A). Looking at fig. 2.4 this would mean that the bound state is simulated from A to B and the free state from B to A. Thus the sum $\Delta G_{\text{bound}}^{\text{A}\rightarrow\text{B}} + \Delta G_{\text{free}}^{\text{B}\rightarrow\text{B}}$ is simulated at once. Since $\Delta G_{\text{free}}^{\text{B}\rightarrow\text{A}} = -\Delta G_{\text{free}}^{\text{A}\rightarrow\text{B}}$ we directly obtain the binding free energy difference (eqn. (2.42)) in a single simulation.

This setup is specifically useful in cases where the mutation introduces or removes a charge and thus there is a charge change in both states. When they are simulated in the same box at the same time the charge will be created in one of the two states and annihilated in the other one and the overall charge of the complete system will not change. The charge distribution of the overall box will still strongly change which can possibly have a strong influence on the water network and the distribution of the surrounding ions. Thus it is likely that the artifacts emerging due to the change of charges in the system will not completely cancel between the legs of the thermodynamic cycle [200]. It could be expected that introducing the changes (changing the λ parameter) slowly enough (“slow growth”) [268] would allow for the complete system to adapt to the change of the charge distribution and thus the resulting binding free energy should converge to the correct result. This assumption is commonly used in the context of free energy pertubation simulations but as shown by Rocklin et al. [200] the artifacts resulting from the use of a discrete solvent and the residual integrated potential components do not cancel between the two legs of the alchemical thermodynamic cycle. Thus deviations from the experimental binding free energy difference should be expected for mutations involving charge changes when using this method. Additionally the standard state corrections (see section 2.5.1) cannot be applied easily since it is not clear which concentrations the free and bound peptide have in these combined simulations boxes.

Simulating bound and free state separately

Instead of simulating both legs of the thermodynamic cycle in a single box together (see section 2.5.1) they can also be simulated separately in a single box each. Thus the two legs cannot interact and mutations introduced in one leg will never influence the other leg and no unknown errors will occur. The resulting mutation free energies can be combined using eqn. (2.42) to obtain the binding free energy difference between states A and B. This setup allows for the application of box size and charge corrections [200, 214, 269–271] for each state. These corrections are applicable to "explicit-solvent simulations employing lattice-sum methods" [200]. The finite-size errors which need to be accounted for as well as their corrections will be explained in section 2.5.1.

Finite-size errors in alchemical free energy explicit-solvent simulations using lattice-sum methods

For systems simulated subject to PBCs the electrostatic interactions are usually evaluated using periodic lattice-sums [8, 9, 200, 204–214]. The most common example is the particle-mesh-Ewald method [8, 9].

Various corrections need to be applied in order to obtain a correct binding free energy from an alchemical simulation. The first is needed in order to obtain comparable free energies from theory, simulation and experiment. Only with respect to a defined standard state the free energies from different computational and experimental methods are comparable [271, 272]. Using the equilibrium constant K_D, the standard binding free energy is defined as

$$\Delta G^0_\mathrm{bind} = -k_\mathrm{B} T \ln(c_0 K_\mathrm{D}), \tag{2.43}$$

where k_B is the Boltzmann constant, T the temperature and c_0 the standard concentration ($1\,\mathrm{mol\,L^{-1}} \approx 1/1661\mathrm{\AA}^{-3}$). In order to obtain the standard binding free energy from a simulation of a finite box the result has to be corrected [271]

$$\Delta G^0_\mathrm{bind} = \Delta G^\mathrm{simulation}_\mathrm{bind} + \Delta G^\mathrm{standardstatecorrection}_\mathrm{bind} \tag{2.44}$$

using the correction term

$$\Delta G^\mathrm{standardstatecorrection}_\mathrm{bind} = -k_\mathrm{B} T \ln\left(c_0 V_\mathrm{simulationbox}\right). \tag{2.45}$$

A second group of corrections is needed to correct the free energies for artifacts resulting from charge changes in a simulation and the treatment of the electrostatic interactions in a finite box. The finite-size error of ligand-charging free energies can be divided into four physical effects, as shown by Rocklin et al. [200], who also introduced a correction scheme for free energy perturbation simulations when using lattice-sum methods where an alchemical transformation of a ligand into a dummy molecule, not interacting with the protein anymore, is performed. In relative alchemical binding free energy simulations a ligand can be transformed into another one or into another state of the same ligand and therefore an adaptation of the correction scheme is needed. This is explained in more detail in section 3.2.3 and the respective supporting information that is shown in chapter 3 and specifically in section A.1.

2.5.2 Endpoint methods

So-called endpoint methods sample the bound and unbound states of a system and then calculate the binding free energy difference between these two states using approximations of the system energy. The most simple method is the linear response approximation or linear interaction energy (LIE) [273, 274]. It is usually applied to obtain protein-ligand free energies and not discussed further here. Popular methods for evaluating protein-protein and protein-ligand binding free energies are the molecular mechanics with Poisson-Boltzmann and surface area solvation (MM/PBSA) [275, 276] and the molecular mechanics generalized Born surface area (MM/GBSA) [275, 276] methods, which will be explained in the next section.

MM/PBSA / MM/GBSA

The MM/PBSA [275, 276] and MM/GBSA [275, 276] methods combine molecular mechanics force fields with continuum solvation models to estimate protein-protein and protein-ligand binding affinities. The free energy of a system state is evaluated using

$$G = E_{\mathrm{bond}} + E_{\mathrm{electrostatic}} + E_{\mathrm{vdW}} + G_{\mathrm{polar}} + G_{\mathrm{non-polar}} - TS \text{ [275–277].} \tag{2.46}$$

The terms E_{bonded} (bond, angle and dihedral energy), $E_{\mathrm{electrostatic}}$ (electrostatic energy), and E_{vdW} (van der Waals interactions) are the corresponding standard molecular mechanics (MM) energy terms. G_{polar} and $G_{\mathrm{non-polar}}$ form the solvation free energy. The distinction between MM/PBSA and MM/GBSA is that the former one uses the Poisson-Boltzmann (PB) equation and the latter the generalized Born (GB) model to estimate the polar solvation term. The non-polar solvation is estimated using the solvent accessible surface area (SASA). The last term includes the system temperature T as well as the entropy S. This is analyzed using normal-mode analysis of the systems vibrational frequencies for a standard state in order to be comparable to experimental values [278]. To obtain binding free energies it is necessary to calculate the free energies of the complex, the unbound ligand and the unbound receptor based on eqn. (2.46) [279]. Commonly only the complex is simulated and the ensemble averages for the free ligand and free receptor are obtained by removing the relevant atoms from the system before the analysis. This leads to

$$\Delta G_{\mathrm{bind}} = \langle G_{\mathrm{receptor+ligand}} - G_{\mathrm{receptor}} - G_{\mathrm{ligand}} \rangle_{\mathrm{receptor+ligand}} \,. \tag{2.47}$$

and to more precise results due to cancellation of intramolecular terms and reduces the required simulation time but ignores the energetic effects of relevant structural changes of the receptor and/or ligand upon binding [277].

Overall, MM/PBSA and MM/GBSA often give better results than LIE, docking and scoring, but worse ones compared to more advanced, e.g. pathway, methods. Depending on the system, reasonably good binding free energy values can be obtained, but for some systems the methods fail. It has been shown that reasonable results can already be achieved using less then 100 repeated simulations with a length of around 200 ps each [277, 280] and therefore much faster than using pathway methods discussed below. The overall coefficient of determination for the whole PDB bind database was shown to be $r^2 = 0.3$ but the individual results differed strongly, $r^2 = 0.0 - 0.8$ [277, 281]. MM/PBSA and MM/GBSA highly depend on the choice of the dielectric constant for the electrostatic energy and the used force field. Additionally, the binding free energy is calculated as a difference of large values and therefore the precision of the result is very low when the standard deviations of the individual terms is high and ligands with similar binding affinities cannot be compared successfully. For a deeper discussion of the MM/PBSA and MM/GBSA methods, their application and possible issues, the reader is referred to [277].

2.5.3 Potential of Mean Force / Pathway Methods

Contrary to the just discussed endpoint methods, free energies of protein-protein binding can also be computed using so-called pathway methods where the free energy is expressed as a function of geometrical reaction coordinates. Such simulations involve a considerably larger computational effort. In most cases, they are more accurate than the former. Most of the following methods are not only used to accelerate binding and unbinding processes but are often also used in simulations of protein folding.

Multiple reviews for these methods and their comparison have been published [282–285].

Thermodynamic integration

One of the oldest pathway methods is the so-called thermodynamic integration (TI) [255] that overcomes barriers in free energy by freezing the chosen CV at different values while sampling along all other degrees of freedom at these fixed points along the reaction coordinate. A free energy profile or PMF is obtained by integrating the mean force, i.e. the derivative of the free energy with respect to the CV [286]. It is possible to slowly move the constraint instead of simulating the CV at fixed values. This is called slow growth [287]. Both methods need comparably long simulation times to reach sufficient sampling along the degrees of freedom of the system and thus a converged PMF. In these methods also the momentum in the direction of the reaction coordinate is constrained and thus they do not fully sample the momentum space.

Umbrella sampling (US)

Umbrella sampling (US) [288] differs from TI by replacing the fixed constraints using restraining biasing potentials, allowing to sample the full momentum space [286]. Here a series of windows is selected along the CV of interest so that the window distributions overlap sufficiently. Usually one uses harmonic potentials of the form

$$w_i(\xi) = \frac{k_i}{2}(\xi - \xi_i^{\mathrm{ref}})^2 \tag{2.48}$$

with center points ξ_i^{ref} and spring constants k_i as biasing restraints for umbrella sampling, but other choices can be imagined. It is also possible to choose an adaptive bias that tries to match the negative of the free energy at each point of the CV ξ. This method is thus called adaptive umbrella sampling [289]. It can be extended to periodically interacting multiple walkers and on-the-fly resampling in order to sample neglected (undersampled) regions. The method is then called adaptive biasing force (ABF) [290] and will be discussed below. If the bias potential is moved or pulled along the CV instead of using a finite number of fixed windows, the method is called steered MD (SMD) or force-probe MD [286, 291]. It will be explained in more detail in the next section.

The most critical part of umbrella sampling is the correct choice of the CVs and of the spring constants k_i defining the strength and thus the width of the biasing potentials. They have to be chosen in advance of the simulations. An advantage of umbrella sampling is that the MD simulations of different windows are completely independent from each other so that they can be executed in parallel. It is even possible to later insert additional windows with larger spring constants if the overlap between originally chosen windows was not sufficient.

Results from umbrella sampling simulations are usually combined either by the umbrella integration [286, 292] or by the popular weighted histogram analysis method (WHAM) [293].

For a more detailed description of umbrella sampling and its analysis the reader is referred to [286].

When it comes to studying the assembly or dissociation of protein complexes, simple umbrella sampling simulations are potentially facing huge sampling problems due to the sheer number of possible relative orientations of the two binding partners, which have to be sampled in every window of the simulation [294]. This is less of a problem if the two proteins each have strongly dipolar character so that they will adopt a preferred orientation relative to each other during association and dissociation processes. The Brownian Dynamics simulations of Spaar et al. discussed above showed that this is the case e.g. for the barnase-barstar complex. When barstar binds to barnase, it approaches barnase “from the right side” due to favorable electrostatic interactions. Also, its binding interface is pre-oriented toward the binding interface like a space-ship that plans to land on the moon. In such cases, it appears plausible to employ a one-directional reaction coordinate to describe protein-protein association and dissociation, e.g. the distance between the proteins’ center of masses. The direction of approach can be taken parallel to the vector connecting the COMs in the bound complex assuming that this is known either from structural studies or from docking. When comparing the unbinding of the three complexes barnase-barstar, cytochrome c – cytochrome c peroxidase, and enzyme 1 – histidine phosphocarrier, it turned out that the PMF computed via umbrella potential simulations had a monotonous uphill profile without transition states. In all cases, the two proteins attracted each other up to distances of about 1.4 to 1.5 nm. Afterwards, the PMF curve was flat reflecting that the protein interaction was shielded by the solvent. Beyond such distances, the relative orientation of the two proteins is not relevant anymore. The same calculations were also performed for a dissociation process starting from a non-specific short-lived contact of the same protein pairs [295]. In that case, the PMF profiles had the same shape, but a shorter attraction basin of only about 0.8 nm and the contact conformation was only about half as stable as the specific complex.

In order to treat sampling issues it is possible to use additional external restraints. Gumbart et al. [26, 294] applied geometrical and conformational restraints that are enforced on the overall relative protein motions and also to constituent amino acids forming the interface region. They showed that the resulting binding free energy of the studied barnase-barstar complex compares well to the experimental value and that the statistical error of the method is low for a system of this complexity. The time the method needs to reach convergence is also much lower than for umbrella sampling without additional restraints. One has to note that the contributions of the external restraints to the overall binding free energy difference were calculated using the adaptive biasing force method. Only the main separation PMF was calculated using umbrella sampling. In a recent study Suh et al. [296] showed that various advanced methods including their newly developed "String Method for Protein-Protein Binding Free-Energy Calculations" can lead to converged results far off from the experimental binding free energy of barnase-barstar. They discuss this in much detail and compare these results to the one obtained by Gumbart et al. [26, 294].

Another recent example of restraint umbrella sampling simulations of protein-protein interactions is found in [297].

Steered MD (SMD)

As already described, steered molecular dynamics (steered MD) or force-probe MD utilizes moving bias potentials in order to push the simulation system over barriers in the free energy. The mean force and thus the PMF can be sufficiently sampled and thus estimated if the movement of the potential is slow compared to the relaxation times of the system [286]. Thus usually very slow movements have to be chosen that drastically increase the amount of sampling in order to converge these simulations. An interesting feature of SMD is its equivalency to atomic-force microscopy [286]. Fast SMD is often used to obtain approximate starting positions for umbrella sampling simulations. If the results of steered MD are evaluated using nonequilibrium analysis methods such as the Jarzynski equality [184] or the Bennet acceptance ratio (BAR) [298] the movement of the potential can be performed much faster than the relaxation time scales of the system. In order to obtain a suitable ensemble average of the PMF these simulations have to be performed repeatedly. A few examples for the usage of steered MD to obtain protein-protein binding free energies are given in [299, 300].

Metadynamics

Metadynamics [301], like all other related methods, strongly depends on the choice of the used reaction coordinates. An advantage of metadynamics is the fact that it is possible to sample multiple CVs at the same time, which is not as easily possible in other methods. Usually only two or three CVs are chosen, because larger numbers would necessitate much more sampling.

In order to overcome barriers in free energy and to escape minima on the free energy surface, metadynamics uses history dependent bias potentials, which are often called hills. These are Gaussian functions dependent on the CVs $\text{¸}(\mathbf{x}) = (\xi_1(\mathbf{x}), \cdots, \xi_n(\mathbf{x}))$ which themselves depend on the coordinates $\mathbf{x}$ of the system. These are added to the potential energy V of the study system at a chosen frequency. The overall resulting bias potential is

$$V_G(\text{¸}(\mathbf{x}), t) = \sum_{t'<t} W(t') \exp\left(-\sum_{i=1}^{n} \frac{(\xi_i(\mathbf{x}) - \xi_i(\mathbf{x}_G(t')))^2}{2\sigma_i^2}\right) \tag{2.49}$$

where $t' = \tau_G, 2\tau_G, 3\tau_G, \cdots$ are multiples of the hill-deposition time τ_G, where $\xi(\mathbf{x}_G(t))$ is the trajectory of the system subject to the action of $V + V_G$, where W is the height of the Gaussian potentials, and where the σ_i are the widths of the Gaussian potentials in the respective CV [301–303]. As is clear from eqn. (2.49) the hills are adding up during the course of the simulation and thus fill up the free energy minima over time. Higher and wider hills increase the speed of convergence but reduce the sharpness of details on the free energy surface.

In metadynamics the free energy surface cannot be calculated from eqn. (2.6) because canonical sampling is hindered by the bias potential and thus it has to be calculated differently. After a certain time the biasing potential has completely filled all minima on the free energy surface and the effective potential becomes flat. At this point convergence is reached. The free energy surface is then simply the negative of the biasing potential

$$G(\xi(\mathbf{x})) = -V_G(\xi(\mathbf{x})) . \tag{2.50}$$

An advancement of the original metadynamics approach is the so-called well-tempered metadynamics [304]. Here the height of the hills is decreased while the bias potential is accumulated, resulting in high hills at the beginning and lower hills at the end of the metadynamics simulations [303], thus strongly improving convergence speed while keeping the sharpness of details on the free energy surface. Furthermore, the well-tempered metadynamics approach leads asymptotically to an exact free energy surface [305], which is highly advantageous over the original approach.

Various extensions to the metadynamics method have been developed since it was first published. These are, for example, funnel metadynamics [306], volume-based metadynamics [307], multiple walkers metadynamics [308], and flux-tempered metadynamics [309], to name just a few [310]. By using a method called infrequent metadynamics [311, 312] it is possible to also calculate the dynamics of a given system, e.g. the kinetic rate constants [310] k_{on} and k_{off}.

Metadynamics in combination with parallel tempering [313] has been successfully used for protein oligomerization/dissociation studies [314, 315].

There are also a range of other methods utilizing history-dependent potentials. One example is the accelerated weight histogram method [316].

For recommendable tutorials on how to perform metadynamics simulations and to choose collective variables the reader is referred to articles describing the tool Plumed [317, 318].

Adaptive biasing force (ABF)

The adaptive biasing force (ABF) method [290] tries to retain the dynamics of the system along the PMF, including the random force described in section 2.5, while additionally leveling the PMF in order to easily move along the PMF, because barriers in free energy are removed. This leads to an acceleration of the passage between the relevant states along the CV [181] and thus improves sampling along the CV. This is achieved by calculating the mean force along a reaction coordinate ξ and removing it via an external biasing force, which is exactly the negative of the current estimate of the mean force. This results in uniform sampling along ξ [182]. ABF has been successfully used to obtain reasonable protein-protein binding free energies, e.g. in [294] and [319]. For a detailed explanation of the ABF method, a suitable choice of the reaction coordinates, error analysis, and extended methods the reader is referred to [181].

Parallel Cascade Selection Molecular Dynamics

Parallel Cascade Selection Molecular Dynamics (PaCS-MD) was first proposed in 2013 by Harada and Kitao as a simulation method to generate conformational transition pathways when reactant and product structures are known *a priori* [320]. In their original manuscript, they state that PaCS-MD differs from other pathway methods like the ones described in the previous sections, because it allows to generate transition pathways without utilizing external perturbations like biasing forces. The method was originally intended to study protein folding and conformational changes but it was extended to protein-ligand (un)binding simulations in 2018 [321]. It was also shown that it is possible to use PaCS-MD to study protein-peptide dissociation [27] and also association and dissociation [322].

The overall framework, as similarly described in [320] and [321] is the following: First an equilibration followed by a short equilibrium MD simulation of the bound state is performed (cycle 0). A certain number n_{rep} of the best snapshots of the resulting trajectory in terms of a previously defined reaction coordinate are extracted from the equilibrium simulation. In this case, best means that these snapshots are the closest to the wanted product structure in terms of the reaction coordinate. For the dissociation PaCS-MD the center-of-mass (COM) distance between the protein and the ligand or peptide is used as a reaction coordinate. The best snapshots are then the ones where the COM distance is largest. From each of these snapshots, a new short MD simulation (further called repeat or replica) with length t_{cyc} is started after a rerandomization of the initial velocities of every atom (cycle 1). Afterwards, all new trajectories are analyzed and again the best n_{rep} snapshots are extracted and used to start new simulations where again the velocities are rerandomized beforehand (cycle 2). This process is repeated (cycles 3, $\cdots$) until the reaction coordinate reaches the value describing the product state or a pre-defined number of cycles is reached. By always selecting the snapshots which are the best in terms of progression along the reaction coordinate the system is brought closer and closer to the product structure while the system is sampled in the intermediate structures without external biases. When the new snapshots are worse in terms of progression along the reaction coordinates, the old snapshots will be kept and the new cycle is started from these. All trajectories are stored for later use for analysis. The challenge in this method is to choose a reasonable number of repeats n_{rep} and the length of the simulations in each cycle t_{cyc} which give reasonable sampling along all important degrees of freedom. It makes sense to perform the whole process a number of times to obtain good sampling, each of these process repeats is called a trial.

Afterwards, the short trajectories can be combined into "reactive trajectories" connecting the initial and final state [320]. In the case of dissociation PaCS-MD these can be used as representative dissociation pathways or initial structures for umbrella sampling [321]. It is also possible to utilize all generated trajectories (also those that do not form the reactive trajectories) to build Markov state models (MSMs, see sec. 2.5.6). Even though a specific reaction coordinate was chosen to rank the snapshots of the cycles of the PaCS-MD simulation no biasing along a reaction coordinate was performed and therefore all positional data for all atoms can be used to build the MSMs. Due to the rerandomization of velocities, the kinetic information is rather less useful. It was shown that it is possible to obtain thermodynamic properties of the system, like the binding or dissociation free energy, by utilizing different features for building the MSMs. Reasonable dissociation free energies were obtained using the inter-COM distance [27, 321], as well as the 3D-COM position of the ligand or peptide relative to the protein [27, 321] and the C_α positions of the peptide relative to the protein [27]. The choice of the cycle length t_{cyc} strongly influences the possibility to choose the lag time $\tau \leq t_{\mathrm{cyc}}$ for the MSM so good thoughts have to be put into the choice beforehand since longer t_{cyc} at a fixed number of replicas n_{rep} extends the total simulation time. A lower number of n_{rep} reduces the possibility of finding good snapshots which show progress along the reaction coordinate.

2.5.4 Replica-exchange methods

In Replica-exchange methods for molecular dynamics [323] the system is simulated in different system states at the same time. At regular time intervals, individual simulations may exchange properties such as temperature or coordinates at certain steps so that barriers in free energy can be overcome more easily. Subsequently, replica-exchange methods were combined with pathway methods to be used in the calculation of free energies and PMFs. A great advantage of many parallel/replica-exchange methods is the possibility of simulating in parallel on multiple computer nodes because the interconnection is only required for the exchange steps for which one does not need high-speed connections between the individual nodes.

Parallel tempering

Increasing the simulation temperature is an obvious way to more easily overcome the free energy barriers along the CVs. Arrhenius law tells us that reaction rates increase with temperature because an increased number of particles have an energy greater than the minimum energy needed for the reaction at increased temperature [183]. Usually simulated tempering methods [324] are performed as follows: First, the system is propagated at a fixed temperature T_i for a number of time steps. Second, the acceptance for switching between two temperatures T_i and T_j is evaluated as a Monte Carlo step with the acceptance probability of

$$\alpha = \min\left(1, \frac{Z_j}{Z_i} \exp\left[-\frac{U(x)}{k_{\mathrm{B}}T_j} + \frac{U(x)}{K_{\mathrm{B}}T_i}\right]\right) \tag{2.51}$$

where i is the index of the present temperature and j the temperature of the new one. In general it is nontrivial to choose the weights Z_i in a suitable way so that all values of i are equivalently sampled [183].

In order to overcome the issue of finding the correct weights it is possible to simulate multiple replicas of the system at the same time at different temperatures. Here not the temperature of a single system is changed but rather the coordinates of two replicas are exchanged with the acceptance probability

$$\alpha = \min\left(1, \exp\left[\left(\frac{1}{k_{\mathrm{B}}T_j} - \frac{1}{k_{\mathrm{B}}T_i}\right)\left(U(x_j) - U(x_i)\right)\right]\right) \tag{2.52}$$

which is independent of the weights Z_i and Z_j. Equal sampling of each index i is achieved by only allowing pairwise swapping [183].

One example of a pathway method extended with parallel tempering is the combination of metadynamics with parallel tempering [313] which has been used in [314, 315] to study protein oligomerization/dissociation. For further information on parallel tempering, the reader is referred to [183, 325].

Generalized/Hamiltonian replica-exchange methods

In the previous section the replicas differed in temperature. It is, however, also possible to change other parameters of the system such as parts of the Hamiltonian [326] or combine various changes, e.g. replicas can have different temperatures and Hamiltonians at the same time and their acceptance probability then reads

$$\alpha = \min\left(1, \frac{\exp\left[-\left(\frac{U_i(x_j)}{k_\mathrm{B}T_i} + \frac{U_j(x_i)}{k_\mathrm{B}T_j}\right)\right]}{\exp\left[-\left(\frac{U_i(x_i)}{k_\mathrm{B}T_i} + \frac{U_j(x_j)}{k_\mathrm{B}T_j}\right)\right]}\right) \text{ [183].} \tag{2.53}$$

2.5.5 Additional pathway methods

The methods explained in the previous sections are just a brief summary of the most popular computational methods to calculate binding free energies of protein-protein interactions. Many more interesting methods, combinations of the aforementioned ones, as well as combinations with machine learning techniques have been successfully used to study association and dissociation of protein-protein systems. It was also beneficial to combine pathway methods with or derive CVs from experimental data such as in SAXS-guided metadynamics [327].

2.5.6 Markov state models

Markov state models (MSMs) have become an important technique for analyzing MD data. They can be used in combination with unbiased MD simulations and various enhanced sampling methods to extract thermodynamics properties and mechanistic insights from the trajectory data. MSMs are models which fulfill the Markov assumption or property, which means that a stochastic process should be memoryless, i. e. the state of the system has no information about the past states and the following state only depends on the present state. This can be summarized in the following equation with $\mathbf{x}, A \in \Omega$ and $\tau \in \mathbb{R}_{0+}$

$$p(\mathbf{x}, A; \tau) = \mathbb{P}[\mathbf{x}(t+\tau) \in A | \mathbf{x}(t) = \mathbf{x}] \tag{2.54}$$

for the probability density p that a trajectory which was started from a certain point $\mathbf{x} \in \Omega$ (the state space Ω usually contains position and velocity when considering MD simulations) at time t will end up in set A at time $t + \tau$ [328]. In biophysics MSMs are often used to analyze data from MD simulations and to extract observables to compare them to experiment.

In order to obtain self-consistent MSMs the states of the system also need to be ergodic and reversible, i. e. for the former condition the network of states must be fully connected [328]. In MD simulation this means that from any state of the system it should be possible to reach any other state, i. e. sufficient sampling is needed, especially when combining multiple simulations with different starting conformations to obtain enough overlap [328]. For the latter condition the probability distribution $p(\mathbf{x}, \mathbf{y}; \tau)$ needs to fulfill *detailed balance* or microscopic reversibility, i. e. for a system in equilibrium, the same number of transitions from $\mathbf{x}$ to $\mathbf{y}$ and from $\mathbf{y}$ to $\mathbf{x}$ need to be observed. Otherwise the model would include source and sink states which leads to a wrong description of the long time scale behavior [328].

The first step when building MSMs from MD data is to project the high-dimensional trajectory data onto a lower-dimensional space by dimensionality reduction onto a feature set, i. e. collective variables, that best catches the variations in the data.

Multiple feature sets can be compared and optimized using the GMRQ [329] or VAMP [330, 331] methods [332]. Common feature sets for binding simulations are protein-ligand (pairwise) distances of amino acids or center of masses. Feature sets usually used to describe internal motions of proteins or for folding simulations are backbone torsions, dihedrals, angles or contacts. The featurized MD data can then be further projected on the time-lagged independent components (tICs) using the time-lagged independent component analysis (tICA) [12–15]. Its goal is to find most time-correlated degrees of freedom existing in the data via linear combinations of features [332]. For tICA a lag time (which is independent of the MSM lag time) has to be chosen to filter out the fastest processes. The result of tICA is that the tICs then display the slowest processes in the simulations. For binding simulations it is likely that the slowest degree of freedom is the binding or unbinding process.

After reducting the dimensionality and projecting the data on to the most time-correlated degrees of freedom the data it has to be decomposed into a number of microstates or clusters. Simulation snapshots should fall into the same microstate, if they are close in an appropriate distance metric and the system can quickly transition between them, i. e. the clustering has to be kinetically relevant. The number of microstates is a weighing between having enough statistics in each state to accurately calculate transition probabilities between states and the goal of having states that best divide the phase space as finely as possible to exclude all large free energy barriers from within states [328]. Poorly sampled microstates can break detailed balance why it makes sense to cut them off after clustering [328]. The most common algorithm to perform the microstate clustering is the k-means [333] algorithm. Usually it is good to use a subsample of the data to cluster it, thereby reducing the impact of outliers [334] and to assign all data to the defined microstates afterwards [328]. Clustering can be difficult for binding simulations due to the different time scales in the bound and unbound states [328] wherefore an eye has to be kept on this while analysing the data.

After clustering, the transition matrix C is built by counting the transitions from and to every state i from every other state j and storing the counts in the matrix entry C_{ij}. For infinite data that would allow to directly calculate the transition probabilities via

$$T_{ij}(\tau) = \frac{C_{ij}}{\sum_k C_{ik}} \text{ [328].} \tag{2.55}$$

With finite sampling and shortcomings in the microstate clustering, this will not work as easily. It has to be assured that detailed balance between any pair of states is fulfilled and a reasonable lag time τ has to be chosen. This choice can be validated using the Chapman-Kolmogorov equation

$$T(n\tau) = T(\tau)^n \tag{2.56}$$

with an integer number of steps n with a length τ each, i. e. taking n steps with a lag time of τ should be equal to taking a MSM with lag time $n\tau$ [328]. This can be used with the relaxation time scales t_i of a model

$$t_i = -\frac{\tau}{\ln \lambda_i} \tag{2.57}$$

with the corresponding eigenvalue λ_i of the transition matrix, because application of the Chapman-Kolmogorov equation will lead to the finding that the relaxation time scale should get nearly constant when increasing τ. The lowest value of τ for which the relaxation time scale gets nearly constant is an appropriate value of τ fulfilling

Markovianity [328].

To obtain models that can be interpreted easily it often makes sense to coarse-grain the MSM into a few macrostates between which the slowest transitions and processes in the system happen. The most common method to combine microstates into macrostates is the "Robust Perron Cluster Cluster Analysis" (PCCA+) spectral clustering method [335–337].

MSMs can be used to obtain rate constants of processes, e. g. the dissociation and association rate constants, k_{off} and k_{on}, of binding and therefore the equilibrium dissociation constant $K_{\text{D}} = \frac{k_{\text{off}}}{k_{\text{on}}}$. One way is to use the transition path theory (TPT) analysis [16].

For a detailed discussion of MSMs, their theoretical background and their applications, the reader is referred to the book by Bowman, Pande and Noé [328].

Chapter 3

How peptide phosphorylation affects its interaction with 14-3-3η domains

This chapter introduces our approaches on understanding the interaction of 14-3-3 proteins with phosphorylated and unphosphorylated peptides. It was already accepted for publishing in PROTEINS: Structure, Function, and Bioinformatics under the title "How phosphorylation of peptides affects their interaction with 14-3-3η domains" [32]. My contributions were the choice of the systems to study and the simulations to perform, performing the simulations, analyzing the data and writing the first draft of the manuscript. In the following sections the preprint version of the manuscript is presented. The abstract is given in the introduction chapter 1.

3.1 Introduction

14-3-3 proteins are a family of adapter proteins performing regulatory functions in cell-cycle control, signal transduction, protein trafficking, and apoptosis. They are abundant in nearly all eukaryotic cells and occur predominantly as stably folded homo- and heterodimers [97, 98] being stabilized by salt bridges and hydrophobic interactions [99, 100]. There exist seven known isoforms in human, namely α/β, γ, ϵ, η, ζ/δ, θ, and σ, exhibiting high sequence and structural similarity (RMS deviations between 0.7 and 1.8 Å) [74]. Each monomer is composed of nine α-helices, forming an amphipathic groove that creates the main binding interface between 14-3-3 and their target proteins [103]. In this groove lysine and arginine residues form a basic cluster that mediates the interaction with 14-3-3 binding partners [101–103].

14-3-3 proteins interact with disordered regions of various phosphorylated proteins that most commonly contain motif 1 (R[S/F/Y/W]XpSXP), motif 2 (RX[S/Y/FW/T/Q/A/D]Xp(S/T)X[P/L/M]) [104–106] or motif 3 (RXXp(S/T)XX-COOH) [105] but also with a few unphosphorylated peptides [101, 103, 107], such as the exoenzyme S (ExoS) [110] or the Carbohydrate-response Element-binding Protein (ChREBP) [108, 109]. To our knowledge there exist hardly any experimental data, such as binding constants, for the unphosphorylated counterparts of phosphorylated motif 1 and motif 2 peptides. The binding of these unphosphorylated peptides seems to be too weak to be measurable [104, 338, 339] and only a lower barrier for the binding constant K_d for the RSRSTSTP peptide exists [104]. Interestingly, it was shown that 14-3-3 dimers are able to bind simultaneously to two target sequences contained in a single polypeptide chain, that leads to an increase in binding affinity compared to the binding at only one binding site [104, 339–341]. This effect has been termed cooperative binding [104]. It was suggested that complexation at the first binding site acts as a gatekeeper necessary for 14-3-3 binding but may not be sufficient in order to enable full biological activity [340].

The main aim of this study was therefore to mechanistically explain why 14-3-3 proteins preferably bind to phosphorylated peptides. To do so we studied complexes of 14-3-3η proteins with phosphorylated motif 1 (c-RAF-pS259 (RSRSTpSTP) and c-Raf-pS233 (HRYpSTP)) as well as motif 2 (RLYHpSLP) peptides together with their unphosphorylated counterparts by means of unbiased molecular dynamics (MD) and alchemical transformation simulations *in silico*.

3.2 Methods

3.2.1 Structure preparation

For this study we selected the 14-3-3η domain with various peptides because their binding affinities have been extensively characterized by Yaffe et al. [104]. So far, no crystal structures for the complex of the 14-3-3η domain bound to the c-RAF-1 peptides RSRSTpSTP, RSRSTSTP, HRYpSTP and HRYSTP have been reported. Thus, we superimposed an X-ray structure of the 14-3-3ζ dimer in complex with the RSRSTpSTPNV peptide (Protein Data Bank (PDB) [6] entry 4FJ3 [341]) onto a crystal structure of the 14-3-3η dimer bound to another peptide (PDB entry 2C63 [342]) using UCSF Chimera [343]. The overall alignment yielded a low RMSD of 0.98 Å for the corresponding C_α atoms. Also the binding modes of the two peptides were very similar (C_α-RMSD with aligned 14-3-3 dimers: 1.52 Å). We used the coordinates of the superimposed RSRSTpSTPNV and HRYpSTP peptides as starting points for the MD simulations of the 14-3-3η dimers bound to these peptides. When studying complexes with the shorter RSRSTpSTP peptide we simply deleted the last two residues.

The same procedure was applied to obtain starting conformations of 14-3-3η bound to the RLYHpSLP peptide. Here, we used an X-ray structure of a 14-3-3ζ dimer binding this peptide (PDB entry 1QJA [344]) and superimposed it onto the same crystal structure of the 14-3-3η dimer binding another peptide (PDB entry 2C63 [342]) as above. Again, the alignment resulted in a very low C_α-RMSD of 0.835 Å. For all monomer simulations we used the structure of the 14-3-3η monomer binding the respective ligand taken from the aligned dimer structures.

In order to obtain starting conformations of 14-3-3η bound to the unphosphorylated peptides we removed the phosphate group from the phosphoserine residues using the rotamer replacement function implemented in UCSF Chimera [343] by applying the serine rotamer from the Dunbrack rotamer library [345]. After energy minimization and equilibration, the structure showed only minor deviations (RSRSTSTP C_α-RMSD of 1.47 Å from the RSRSTpSTP structure in the bound state).

For the alchemical free energy simulations, hybrid structures and topologies including the unphosphorylated and phosphorylated residues in the same file were obtained using the tool PMX [267, 346]. Since no hybrid structures for phosphorylated serines existed in the published program, these were kindly provided by V. Gapsys for the CHARMM36m [192] force field.

3.2.2 Molecular dynamics simulations

All MD simulations were performed using the GROMACS [4] 2018.8 software package with the CHARMM36m [192] force field for proteins and peptides. Water molecules were represented by the TIP3P [224] model modified for the CHARMM force field [190]. Hydrogen atoms were generated using the internal GROMACS tool pdb2gmx. In the simulations we applied a time step of 2 fs and periodic boundary conditions [201]. Long-range electrostatic interactions were computed with the particle-mesh-Ewald summation method [8] and the non-bonded interaction cutoff was set to 12 Å. All bonds were constrained using the LINCS [347] algorithm (in plain MD simulations only H-bonds were constrained). Long range dispersion corrections were applied for energy and pressure. A concentration of 0.15 mol L^{-1} NaCl was included in all simulation boxes in order to mimic physiological conditions.

Initially, the systems were minimized in two steps. First the unsolvated molecular system was minimized using the steepest descent algorithm for a total of 50000 steps with an initial step size of 0.01 nm and a convergence value for the maximal force of 5 $\text{kJ mol}^{-1}\,\text{nm}^{-1}$. After solvating and ionizing the system, a minimization was carried out, using the steepest descent algorithm for a total of 100000 steps with an initial step size of 0.01 nm and a convergence value of 500 $\text{kJ mol}^{-1}\,\text{nm}^{-1}$. During these steps the positions of the heavy protein and peptide atoms were kept rigid using position restraints.

Afterwards the systems were thermalized for 500 ps each at 100 K, 200 K, and at the final temperature of 298.15 K, respectively while keeping the position restraints for the proteins and peptides. These steps used the velocity-rescaling thermostat [239] with a coupling time constant of 0.1 ps and a separate temperature bath for solute and solvent. The equilibration was continued for another 500 ps in the NPT ensemble by adding a Berendsen barostat [238] with a time constant of 2.0 ps, a reference pressure of 1 bar, and an isothermal compressibility of $4.6 \cdot 10 \times 10^{-5}\,\text{bar}^{-1}$ while still keeping the position restraints of the solute. Thereafter, the Berendsen barostat [238] was changed to a Parrinello-Rahman extended-ensemble pressure coupling [250, 348] because the former does not yield a correct thermodynamic ensemble, but is very efficient for the scaling of a box during the start of a simulation. The position restraints were released in three steps from 1000 via 100, and 10 to 0 kJ/mol/nm^2.

Subsequently, unbiased production runs were performed. The trajectory was recorded every 10 ps.

Unbiased MD simulations

For the plain MD simulations a triclinic box was created such that the minimal distance of every solute atom from the edge of the box was at least 1.5 nm. This resulted in a size of 8.5 nm × 7.5 nm × 9.0 nm for the 14-3-3η monomer and 12.0 nm × 8.0 nm × 10.0 nm for the 14-3-3η dimer simulations, respectively.

We simulated two systems in the absence of a peptide: the 14-3-3η monomer and the 14-3-3η dimer, as well as six systems containing a single peptide: the 14-3-3η monomer with the peptides RSRSTpSTP, RSRSTSTP, RLYHpSLP, and RLYHSLP and the 14-3-3η dimer with the peptides RSRSTpSTP and RSRSTSTP. Additionally, we simulated another four dimer systems where each monomer was bound to a peptide: the 14-3-3η dimer with RSRSTpSTP and HRYpSTP, with RSRSTpSTP and HRYSTP, with RSRSTSTP and HRYpSTP and with RSRSTSTP and HRYSTP. Each system was simulated three times for 1 µs each. To compare the results with the free states we also simulated the unbound RSRSTSTP/RSRSTpSTP and RLYHSLP/RLYHpSLP peptides in a water box, each one once for 1 µs. All together this sums up to a total of 40 µs of simulation time.

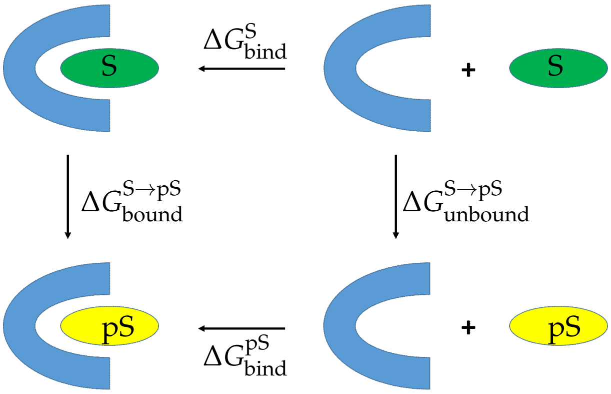


FIGURE 3.1: Thermodynamic cycle to compute the difference in binding free energy between phosphorylated (denoted pS) and unphosphorylated peptides (on serine, denoted S) complexed in complex with a protein domain (depicted in blue). The figure was taken from ref. [32].

Analysis tools

Hydrogen bonds and contacts were identified in the trajectories using the MDTraj[349] python toolkit using the "Baker-Hubbard" [350] and "compute contacts" algorithms. The number of hydrogen bonds formed by the phosphoserine or serine residues at the relevant positions were computed with the gmx hbond tool. The Root mean square fluctuations (RMSFs) were characterized by the gmx rmsf tool and the principal component analysis (PCA) was carried out with the gmx colvar and gmx aneig tools, all included in GROMACS [4].

3.2.3 Alchemical free energy simulations

Alchemical free energy differences were computed by performing repeated non-equilibrium simulations along the reaction path and calculating the free energy difference using two methods based on the Jarzynski equality[184], namely the Bennett acceptance ratio (BAR) [298] and the Crooks Gaussian Intersection [351] method, respectively.

The four relevant states of each system were the unbound phosphorylated peptide, the unbound unphosphorylated peptide, the bound phosphorylated peptide and the bound unphosphorylated peptide. These can be arranged into the thermodynamic cycle shown in Fig. 3.1. The difference in binding free energy is then obtained from the thermodynamic cycle as

$$\Delta\Delta G_{\mathrm{bind}} = \Delta G^{\mathrm{pS}}_{\mathrm{bind}} - \Delta G^{\mathrm{S}}_{\mathrm{bind}} = \Delta G^{\mathrm{S\to pS}}_{\mathrm{bound}} - \Delta G^{\mathrm{S\to pS}}_{\mathrm{unbound}} \,. \tag{3.1}$$

The superscript S denotes the unphosphorylated peptide containing a serine while the superscript pS denotes the phosphorylated peptide containing a phosphoserine. These kind of simulations do not allow to calculate absolute free energy values for the binding of a peptide to a protein but yield the difference in binding free energy between two different peptides (here, the phosphorylated and unphosphorylated peptide) binding to the same protein. In free energy perturbation methods the change from one state A to the other state B is typically achieved by using a coupling parameter λ that changes the topology from A to B gradually. During the course of the simulation the work performed on the system is computed and the change in free energy during that process is evaluated using the Bennett acceptance ratio (BAR) [298]. In our case the resulting binding free energy difference is calculated as the difference between the two alchemical transformations in the bound and free state (see eqn. (3.1)).

In experiments one usually measures the dissociation constant K_D that is directly related to the binding free energy by

$$\Delta G^0_\mathrm{bind} = -k_\mathrm{B}T\ln(c_0K_\mathrm{D}), \tag{3.2}$$

with the Boltzmann constant k_B, temperature T, at the standard concentration $c_0 = 1\,\mathrm{mol\,L^{-1}} \approx 1/1661\ \mathrm{\mathring{A}}^{-3}$. Since the transformations of the free and bound peptides are performed in simulation boxes of different, finite sizes resulting in different concentrations of the solute, both transformation free energies have to be corrected for the box size using the standard concentration. The corrected standard binding free energy [269, 271] is then

$$\Delta G^0_\mathrm{bind} = \Delta G^\mathrm{simulation}_\mathrm{bind} - k_\mathrm{B}T\ln(c_0V_\mathrm{box}) \tag{3.3}$$

whereby V_box is the volume of the simulation box.

Introducing a phosphate group during the course of a simulation substantially alters the total charge of the system. This can lead to artifacts in the simulations and thus the calculated free energies have to be corrected [200, 213, 214, 352, 353]. An elaborate correction scheme when using lattice-sum methods was derived by Rocklin et al. [200]. This method was originally designed for free energy perturbation simulations calculating absolute binding free energies where the bound ligand is alchemically transformed into a non-interacting dummy molecule. In our alchemical transformations the ligand is transformed from one state to another one and thus the correction scheme has to be adapted. Chen et al. [354] derived a correction scheme for relative binding free energies for ligands with different charges, as is the case in our systems. They assumed that the protein charge is zero, thus all terms involving it were left out. We also included the empirical correction term of Rocklin et al. [200]. A derivation is shown in the appendix. The resulting corrections terms are

$$\Delta\Delta G^{\mathrm{A}\to\mathrm{B}}_{\mathrm{NET\_USV,\,state}}(L_\mathrm{state}) = -\frac{\xi_{LS}}{8\pi\epsilon_0\epsilon_S}\left[(Q_L^\mathrm{B})^2 - (Q_L^\mathrm{A})^2 + 2Q_P(Q_L^\mathrm{B} - Q_L^\mathrm{A})\right]\frac{1}{L_\mathrm{state}} \tag{3.4}$$

for the periodicity-induced net-charge interactions and undersolvation [200],

$$\Delta\Delta G^{\mathrm{A}\to\mathrm{B}}_{\mathrm{RIP,\,state}}(L_\mathrm{state}) = \left[(I^\mathrm{B}_{P,\mathrm{state}} + I^\mathrm{B}_{L,\mathrm{state}})Q^\mathrm{B}_L - (I^\mathrm{A}_{P,\mathrm{state}} + I^\mathrm{A}_{L,\mathrm{state}})Q^\mathrm{A}_L + (I^\mathrm{B}_{L,\mathrm{state}} - I^\mathrm{A}_{L,\mathrm{state}})Q_P\right]\frac{1}{L^3_\mathrm{state}} \tag{3.5}$$

for the residual integrated potential effects [200],

$$\begin{aligned}\Delta\Delta G^{\mathrm{A}\to\mathrm{B}}_{\mathrm{EMP,\,state}}(L_\mathrm{state}) =& -\frac{1}{8\pi\epsilon_0}\frac{16\pi^2}{45}\left(1-\frac{1}{\epsilon_S}\right)\\ &\times\left(\left[(Q_P + Q_L^\mathrm{B})^2 - Q_P^2\right](R^\mathrm{B}_{L,\mathrm{state}})^5 - \left[(Q_P + Q_L^\mathrm{A})^2 - Q_P^2\right](R^\mathrm{A}_{L,\mathrm{state}})^5\right)\frac{1}{L^6_\mathrm{state}}\end{aligned} \tag{3.6}$$

as additional empirical correction term [200] with $R^{\mathrm{X}}_{L,\mathrm{state}} = \left(\left[\frac{1}{8\pi\epsilon_0} \frac{4\pi}{3} \left(1 - \frac{1}{\epsilon_S}\right) Q^{\mathrm{X}}_L \right]^{-1} I^{\mathrm{X}}_{L,SLV,\mathrm{state}} \right)^{1/2}$ and

$$\Delta\Delta G^{\mathrm{A}\to\mathrm{B}}_{\mathrm{DSC,\,state}}(L_{\mathrm{state}})(N_S, L) = -\frac{\gamma_S}{6\epsilon_0} \frac{N_{S,\mathrm{state}}}{L^3_{\mathrm{state}}} (Q^{\mathrm{B}}_L - Q^{\mathrm{A}}_L) \tag{3.7}$$

for discrete solvent effects. "X" can be either state "A" or "B", Q_P is the net charge of the protein, Q^{X}_L denotes the net charge of the ligands, ϵ_S is the static relative dielectric permittivity of the solvent, $\xi_{LS} \approx -2.837297$ is the cubic lattice-sum (Wigner) integration constant, "state" can be either "bound" or "free", and L_{state} is the size of the simulation box in the respective state.

For the alchemical simulations of a 14-3-3η monomer bound to a peptide and another unbound peptide, a cubic box was created such that the minimal distance of every atom of the solute was at least 1.5 nm from the box edge resulting in a size of around 10.2 nm per edge for the simulations of the bound state and 5.3 nm per edge for the simulations of the free state. The (unbiased) equilibrium simulation lasted 20 ns. The first 4 ns were discarded and 100 starting structures for the alchemical transformations were recorded every 16 snapshots (i. e. every 160 ps). Each of these 100 transformations was performed for 1 ns by gradually changing the parameter λ each step by $\Delta\lambda = 0.2 \cdot 10^{-5}$ from 0 to 1 during the course of the simulation. The change of the Hamiltonian with respect to the coupling parameter λ was saved every time step. We repeated these simulations four to seven times for every system to obtain a statistical error.

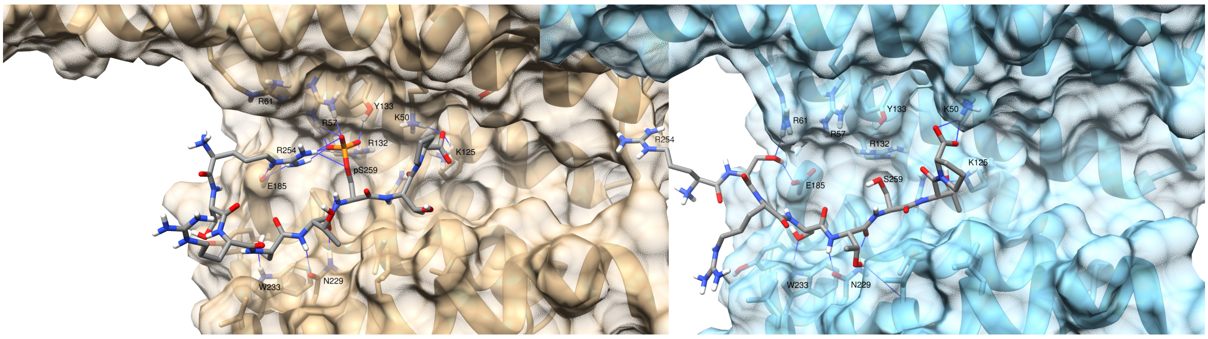
To ensure smooth convergence of the derivatives of the non-bonded energy terms while creating or annihilating atoms, we used a soft-core potential with soft-core parameter $\alpha_{\mathrm{SC}} = 0.3$, soft-core power $p_{\mathrm{SC}} = 1$ and radius of the interaction set to $\sigma_{\mathrm{SC}} = 0.25$. The alchemical transformations were performed both in the forward and backward directions, i.e. from $\lambda = 0$ to 1 and from $\lambda = 1$ to 0, in order to achieve better convergence. The results were then combined using an analysis tool included in the PMX package [267], which calculates the free energy difference using the Bennett acceptance ratio (BAR) [298], the Jarzynski equality[184] and the Crooks Gaussian Intersection [351] method. All results shown in this study were obtained using BAR, but results were only considered valid when all three estimators gave similar estimates, which is a good measure for convergence of such non-equilibrium methods [184, 355].

As stated before, the simulations for the transformation of the free and bound peptides were performed in simulation boxes of different, finite sizes resulting in different concentrations of the solute. Thus the results were corrected for the box size using the standard concentration c_0 [271] and likewise for electrostatic finite-size effects [200].

3.3 Results

3.3.1 Unbiased MD simulations

MD simulations of 14-3-3η monomers and dimers complexed with either phosphorylated or unphosphorylated peptides or both were stable on the 1 µs time scale studied here. The mean C_α-RMS deviations from (modified) crystal structures are listed in table 3.1. The only exceptions to this were a single simulation of 14-3-3η with the RLYHSLP peptide and another simulation of a 14-3-3η dimer with the RSRSTpSTP and HRYSTP peptides bound at the same time. In these simulations unbinding of the unphosphorylated peptides happened. Hence, we repeated the one



(A) RSRSTpSTP peptide (SP2) bound to 14-3-3η (B) RSRSTSTP peptide bound to 14-3-3η

FIGURE 3.2: Representative structures of a phosphorylated RSRSTpSTP (doubly charged, SP2; left hand side) and a unphosphorylated RSRSTSTP peptide (right hand side) bound to 14-3-3η obtained by a centroid search of all trajectory frames. Pairwise RMSD was computed as a distance metric and the pairwise distances were then used to calculate a pairwise similarity. The centroid is the frame with the highest sum of similarities. The figures were taken from ref. [32].

	Simulation	**mean C_α-RMSD [Å]**
14-3-3η monomer	without peptide	2.952 ± 0.001
	with RSRSTSTP	3.686 ± 0.002
	with RSRSTpSTP	2.909 ± 0.001
	with RLYHSLP	3.355 ± 0.001
	with RLYHpSLP	2.795 ± 0.001
14-3-3η dimer	without peptides	3.628 ± 0.002
	with RSRSTSTP	3.397 ± 0.002
	with RSRSTpSTP	3.794 ± 0.002
	with RSRSTSTP + HRYSTP	3.995 ± 0.002
	with RSRSTSTP + HRYpSTP	3.566 ± 0.002
	with RSRSTpSTP + HRYSTP	4.127 ± 0.003
	with RSRSTpSTP + HRYpSTP	3.490 ± 0.001

TABLE 3.1: Mean C_α-RMSD from starting structures. The table was taken from ref. [32].

RSRSTpSTP (31 contacts)

Residue pair	Distance	Int. energy
ARG57 – SP2259	1.7	-512.26
ARG132 – SP2259	1.66	-488.15
LYS50 – PRO261	2.26	-250.97
LYS125 – PRO261	3.14	-250.01
LYS50 – SP2259	3.5	-240.85
GLU185 – ARG254	3.3	-233.96
TYR133 – SP2259	1.7	-120.62
LYS125 – THR260	2.49	-48.92
ASN178 – THR260	1.93	-40.68
LYS50 – THR260	3.74	-30.07
ASN229 – THR258	2.09	-26.34
ASN178 – SP2259	2.58	-25.64
ASN229 – SER257	2.64	-19.68
VAL181 – SP2259	2.5	-18.23
TRP233 – SER257	2.57	-14.28
LEU232 – ARG256	2.77	-7.32
LEU225 – SP2259	3.42	-5.66
LEU225 – THR258	2.44	-5.23
GLY174 – THR260	2.36	-5.19
ILE222 – THR260	2.57	-4.89
LEU177 – SP2259	2.86	-4.83
GLU185 – SER257	3.73	-4.8
LEU177 – THR260	2.33	-3.59
TYR184 – SER257	2.84	-3.56
VAL181 – THR258	3.35	-3.08
LEU177 – THR258	2.93	-2.49
VAL181 – SER257	3.11	-2.46
LEU225 – THR260	2.76	-0.88
LEU232 – SER257	2.71	-0.79
ARG132 – ARG254	3.9	0.0
ARG57 – ARG254	3.34	123.28

RSRSTSTP (13 contacts)

Residue pair	Distance	Int. energy
LYS50 – PRO261	2.07	-239.55
LYS125 – PRO261	3.21	-211.84
LYS125 – THR260	2.39	-56.84
ASN178 – THR260	2.47	-23.84
ASN178 – SER259	2.71	-18.34
ARG132 – SER259	3.96	-13.34
LEU225 – THR258	3.94	-5.74
LEU177 – SER259	2.56	-3.48
LEU177 – THR260	3.17	-3.19
ILE222 – THR260	3.93	-2.41
LEU177 – THR258	3.51	-2.11
GLY174 – THR260	3.71	-1.47
VAL181 – SER259	3.56	-1.17

TABLE 3.2: Residue contacts between a 14-3-3η monomer and RSRSTpSTP/RSRSTSTP peptides and respective mean interaction energies in kJ/mol. Here, all contacts were counted if the mean of the shortest atom distance (in Å) between two residues over all frames of three repeated simulations was below 4 Å. The first residue belongs to the protein and the second one to the peptide. SP2259 denotes the phosphorylated (doubly charged) SER259. The tables were taken from ref. [32].

with RLYSLP and cut the trajectory with RSRSTpSTP and HRYSTP in order to keep only the part of the trajectory without the unbinding event in order to obtain trajectories without unbinding events. Fig. 3.2 shows representative snapshots from the simulations of a 14-3-3η monomer with the RSRSTpSTP and RSRSTSTP peptides. Both peptides remained bound in the binding groove.

The X-ray structure of 14-3-3η bound to RSRSTpSTP and the conformations sampled in the simulations suggest that there exists a strong difference in how tightly phosphorylated and unphosphorylated peptides are coordinated in the 14-3-3 binding groove. To quantify this, we extracted mean distances between peptide and protein residues from the plain MD simulations. Protein and peptide residues were defined to be in contact when the mean distance between them was below 4 Å over the full simulation length. Overall, the 14-3-3η protein formed a total of 31 relevant contacts with the RSRSTpSTP peptide (four contacts had distances of less than 2 Å) and 13 relevant contacts with the RSRSTSTP peptide (no contact had a distance shorter than 2 Å) (see table 3.2).

In particular, both peptides formed very strong interactions between their Pro261 and Lys50 and Lys125 of the protein (van-der-Waals contacts and very negative mean interaction energies) even though hydrogen bonds between these residues are only found in around $35-40\%$ of the frames with the phosphorylated RSRSTpSTP peptide bound and in $65-70\%$ of the frames with the unphosphorylated RSRSTSTP peptide bound (see table 3.4). Similar interactions were observed for the 14-3-3η monomer binding RLYHSLP/RLYHpSLP with hydrogen bonds and contacts of Pro9 of the peptide with Lys50 and Lys125 of the protein and hydrogen-bond occupancies in the same range (see table A.1), as well as for 14-3-3η dimers binding two peptides (RSRSTSTP/RSRSTpSTP + HRYSTP/HRYpSTP) at the same time. For the 14-3-3η dimer complexed with RSRSTpSTP no hydrogen bonds were formed that involve Pro261 whereas the occupancy of the hydrogen bonds for Pro261 for RSRSTSTP was much lower for the dimer than for the monomer bound to RSRSTSTP (see table A.11). In both cases Pro261 formed contacts with Ser46 and Val47 of the protein (see table A.6). These interactions likely exist due to the negatively charged C-terminal placed at Pro261 (for RSRSTSTP/RSRSTpSTP) or Pro9 (for RLYHSLP/RLYHpSLP) in the simulated peptides. This could be an artifact of simulating only a short peptide instead of a long one, which would place its C-terminal outside of the binding groove. This could also be the reason why we observed nearly no unbinding of the unphosphorylated peptides in our simulations even though their binding affinity to 14-3-3 proteins is too weak to be measured in experiment [104, 338, 339]. Thus the binding affinity of the unphosphorylated peptides is likely altered and possibly overestimated in our later performed alchemical simulations of 14-3-3η monomers with the RSRSTSTP/RSRSTpSTP and RLYHSLP/RLYHpSLP peptides. Previously it was argued that a proline at position +2 from the phosphoserine opens up the possibility for the peptide chain to make a turn in order to be able to also bind to the other binding pocket of a 14-3-3 dimer [104–106].

We also compared the number of hydrogen bonds accepted by the phosphate oxygen atoms of phosphoserine in the bound and free (completely solvated) states (see table 3.3). The phosphate of the phosphorylated RSRSTpSTP peptide formed slightly more hydrogen bonds in the bound state compared to the free state. This holds true no matter if the OG oxygen from the original serine side chain is taken into account ($\Delta\# = 0.109 \pm 0.017$, $d = 0.08$) or not ($\Delta\# = 0.453 \pm 0.018$, $d = 0.40$), but calculating Cohen's d shows that the former is not a relevant difference whereas the latter one is. If only the hydrogen bonds involving the oxygen atoms and the protein or water are considered, instead of also including intra-peptide hydrogen

RSRSTpSTP

	H_2O + protein + peptide		**H_2O + protein**	
atom	**bound**	**free**	**bound**	**free**
OG	0.395 ± 0.002 (SD: 0.547)	0.740 ± 0.002 (SD: 0.640)	0.240 ± 0.002 (SD: 0.451)	0.740 ± 0.002 (SD: 0.640)
OE1	3.580 ± 0.003 (SD: 0.576)	3.426 ± 0.003 (SD: 0.770)	2.539 ± 0.003 (SD: 0.692)	1.881 ± 0.004 (SD: 0.992)
OE2	3.122 ± 0.002 (SD: 0.346)	3.402 ± 0.003 (SD: 0.730)	3.122 ± 0.002 (SD: 0.346)	1.826 ± 0.004 (SD: 1.039)
OE	4.041 ± 0.003 (SD: 0.674)	3.461 ± 0.003 (SD: 0.735)	3.413 ± 0.003 (SD: 0.676)	3.461 ± 0.003 (SD: 0.735)
Total (no OG)	10.743 ± 0.006 (SD: 0.952)	10.289 ± 0.008 (SD: 1.294)	9.074 ± 0.006 (SD: 1.027)	7.168 ± 0.009 (SD: 1.613)
Total	11.137 ± 0.008 (SD: 1.098)	11.029 ± 0.010 (SD: 1.444)	9.314 ± 0.007 (SD: 1.122)	7.907 ± 0.011 (SD: 1.736)

RSRSTSTP

	H_2O + protein + peptide		**H_2O + protein**	
atom	**bound**	**free**	**bound**	**free**
OG	0.978 ± 0.003 (SD: 0.627)	1.191 ± 0.002 (SD: 0.623)	0.975 ± 0.003 (SD: 0.628)	1.173 ± 0.002 (SD: 0.629)

TABLE 3.3: Mean number (#) of hydrogen bonds of each oxygen atom in the serine/phosphoserine side chains of the RSRSTSTP/RSRSTpSTP peptides in the bound (monomer simulation) and free state. Columns two and three termed H_2O + protein + peptide (free: H_2O + peptide) list the counts of hydrogen bonds between the oxygen atoms of (phospho)serine listed in the first column from the left and atoms from the surrounding water, the protein and the rest of the peptide (free: water and peptide). Columns four and five labeled H_2O + protein (free: H_2O) contain the counts of hydrogen bonds without the ones connecting the oxygen atoms to the peptide itself. The tables were taken from ref. [32].

RSRSTpSTP (26 pairs)

Residue pair	Occupancy [%]
TYR133-OH – SP2259-OE2	99.99
ARG132-NH1 – SP2259-OE1	99.91
ARG132-NH2 – SP2259-OE2	99.17
ASN178-OD1 – THR260-N	98.23
ARG57-NH2 – SP2259-OE1	97.31
ARG57-NH1 – SP2259-OE	96.11
ASN229-ND2 – THR258-O	83.71
ASN178-ND2 – THR260-O	66.00
ASN229-OD1 – THR258-N	48.12
TRP233-NE1 – SER257-OG	39.83
LYS50-NZ – PRO261-OXT	39.25
LYS50-NZ – PRO261-O	37.74
LYS125-NZ – PRO261-OXT	37.24
LYS125-NZ – PRO261-O	34.88
ARG57-NH1 – SP2259-OE1	34.41
GLU185-OE1 – ARG254-NH2	31.68
GLU185-OE2 – ARG254-NH2	31.09
LYS125-NZ – THR260-OG1	29.84
LYS125-NZ – THR260-O	28.53
ASN178-OD1 – THR260-OG1	27.03
ARG132-NH1 – SP2259-OE2	26.80
GLU185-OE2 – ARG254-NE	23.47
GLU185-OE1 – ARG254-NE	21.38
LYS50-NZ – SP2259-OE	18.23
ASN178-ND2 – THR260-OG1	15.58
ASN229-OD1 – SER257-OG	11.10

RSRSTSTP (20 pairs)

Residue pair	Occupancy [%]
ASN229-ND2 – THR258-O	86.86
ASN229-OD1 – THR258-N	75.79
ASN178-OD1 – THR260-N	74.06
LYS50-NZ – PRO261-OXT	69.97
LYS50-NZ – PRO261-O	66.16
TRP233-NE1 – SER257-OG	43.83
LYS125-NZ – THR260-OG1	39.35
ASN178-ND2 – THR260-O	38.32
ASN178-ND2 – THR260-OG1	37.01
GLU185-OE1 – SER255-OG	28.99
ARG132-NH1 – SER259-OG	28.19
LYS125-NZ – PRO261-O	24.68
GLU185-OE2 – SER255-OG	23.05
GLU185-OE2 – SER257-OG	20.75
GLU185-OE2 – SER257-N	20.63
LYS125-NZ – PRO261-OXT	17.08
GLU185-OE2 – ARG256-N	16.44
GLU185-OE1 – SER257-OG	13.80
GLU185-OE1 – SER257-N	13.53
LYS125-NZ – THR260-O	13.08

TABLE 3.4: Percent of hydrogen-bond occupancies for hydrogen bonds between 14-3-3η monomers and RSRSTSTP/RSRSTpSTP peptides, which exist in more than 10 % of all simulation frames. The first residue belongs to the protein and the second one to the peptide. The tables were taken from ref. [32].

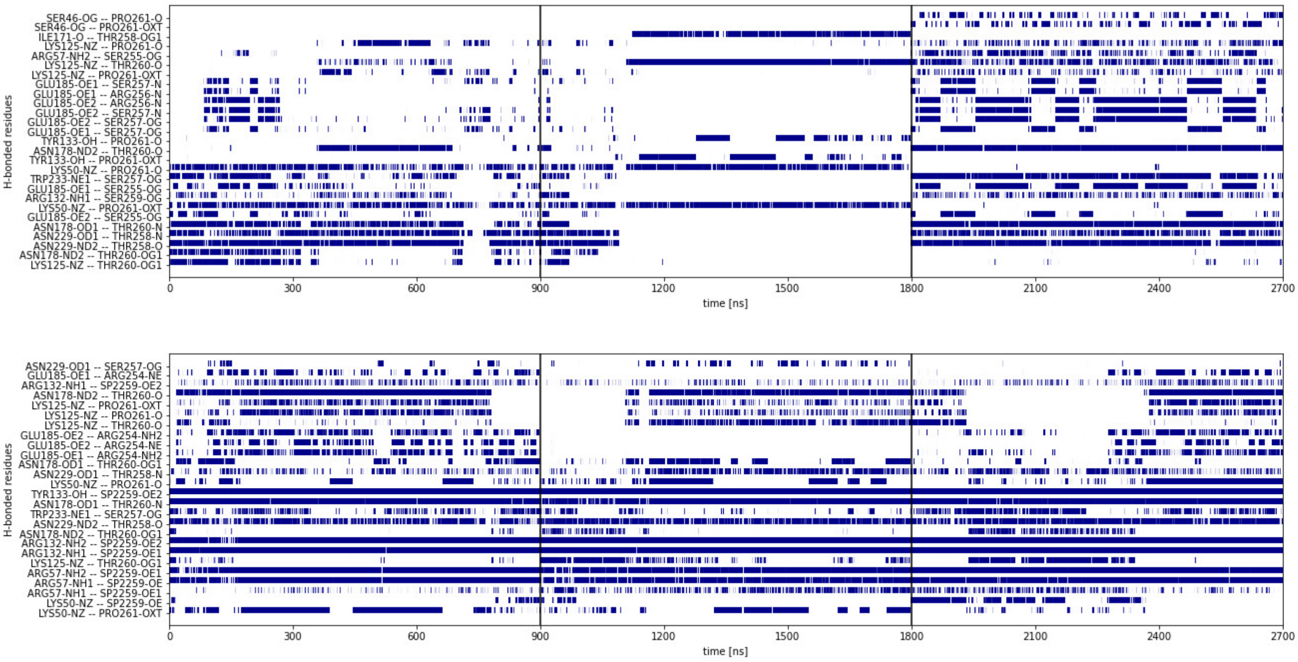


FIGURE 3.3: Hydrogen bonds present in more than 10 % of all frames (white = no hydrogen bond, blue = hydrogen bond) between a 14-3-3η monomer and the RSRSTSTP (top) and RSRSTpSTP (bottom) peptides in a monomer simulation. The figure was taken from ref. [32].

bonds, the effect is even more pronounced ($\Delta\# = 1.406 \pm 0.019$, $d = 0.96$ with OG and $\Delta\# = 1.906 \pm 0.015$, $d = 1.41$ without OG). Thus, the phosphate is energetically better coordinated in the bound state than in the free state. Also three of the four very close contacts (< 2 Å) between the protein and the RSRSTpSTP peptide are formed as strong salt bridges between a charged protein residue and the phosphoserine.

For the unphosphorylated RSRSTSTP peptide one can only analyze the OG atom of the serine side chain. Its coordination in the free state is clearly stronger than in the bound state. Also, its hydrogen-bonding pattern shows more variability (highest occupancy is around 87 %) than for the phosphorylated peptide where multiple hydrogen bonds involving mostly phosphoserine exist for close to 100 % of the time (see table 3.4). The same holds true for the 14-3-3η dimer binding a single RSRSTSTP/RSRSTpSTP peptide (see table A.3). The representative snapshot of the bound RSRSTSTP peptide in Fig. 3.2 also supports these findings. In that figure, the serine side chain is trying to orient towards the surrounding water instead of binding to a 14-3-3η residue. The same also applies to the 14-3-3η dimer binding the RSRSTSTP and RSRSTpSTP peptides (see tables A.3, A.4 and A.5). Interestingly, the interaction between Asp178 of the protein and the backbone of peptide residue Thr260 is much tighter and more stable when the phosphorylated peptide rather than the unphosphorylated RSRSTSTP peptide is bound either to the 14-3-3η monomer (see tables 3.2 and 3.4) or to the 14-3-3η dimer (compare tables A.6 and A.3). The stronger contacts formed by phosphorylated peptides are thus not only due to the direct interactions of the phosphate with charged protein residues but also involve further hydrogen bonds/contacts.

When the 14-3-3η monomer is instead bound to the RLYHSLP and RLYHpSLP

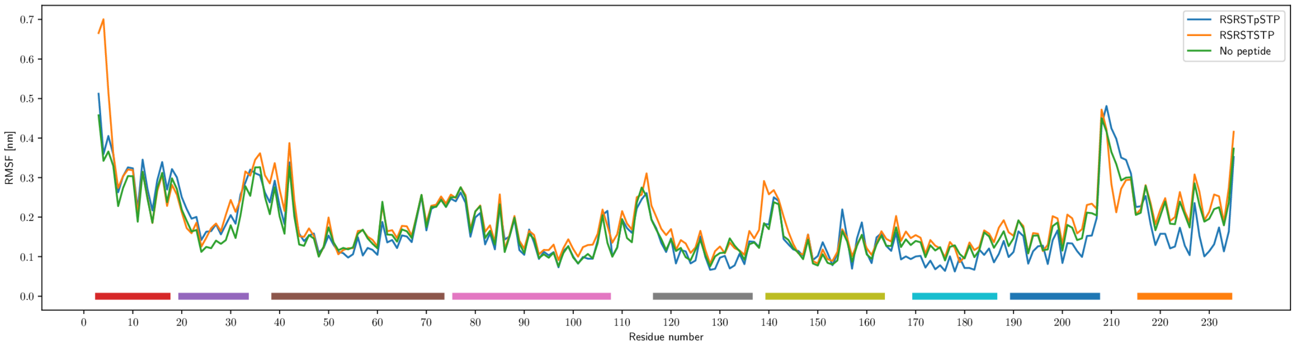


FIGURE 3.4: Root mean square fluctuations (RMSFs) of the residues in the free 14-3-3η monomer and when bound to the RSRSTSTP and RSRSTpSTP peptides. The helices of the monomer are indicated as colored bars at the bottom of the plot. The figure was taken from ref. [32].

peptides, one observes a very similar coordination of the phosphate group in the bound and free states (the effect is much stronger here, see table A.1). Yet the hydrogen bonds are more variable for the phosphorylated RLYHpSLP peptide than for the motif 1 peptide RSRSTpSTP (see table A.2). A particularly noteworthy interaction is that between Asp229 of the 14-3-3η monomer and His6 of the peptide, which is at position +1 from the serine/phosphoserine. This tight interaction seems to pull the entire peptide chain away from the basic pocket that usually accommodates the phosphate group, as well as the side chain of the serine. This effect explains the reduced hydrogen-bond occupancy of the phosphate group but could also further reduce the binding affinity of the unphosphorylated peptide, because all interactions with the binding pocket are reduced in favor of this single interaction.

Next we investigated the contacts between peptide residues next to phosphoserine and protein residues. Interestingly, such contacts mostly involved backbone atoms of these peptide residues (except for His6 in the RLYHSLP/RLYHpSLP peptides). This matches the experimental findings that motif 1 and 2 peptides do not have preferred amino acids at positions -1 and +1 next to the phosphorylated residue.

In the simulation of the 14-3-3η monomer with phosphorylated RSRSTpSTP peptide and the one of the dimer with the same peptide we noticed interactions between Arg254 and Arg256 of the peptide and the phosphate group of pSer259 (combined in around 58 % of frames). Similar interactions were formed between Arg3 and the phosphate of pSer7 in the simulations of a 14-3-3η monomer with the phosphorylated RLYHSLP peptide (in around 78 % of frames). These interactions led to a nearly completely closed positively charged pocket that shields the phosphoserine from the water and leads to an even better coordination of the phosphate oxygen atoms. No such interaction with the respective serine side chain was found in simulations where an unphosphorylated peptide was bound to 14-3-3η. In order to study how peptide binding affects the stability of the 14-3-3η monomer we then performed an RMSF analysis of the trajectories. Fig. 3.4 shows that the binding of peptides did not notably influence the fluctuations of the 14-3-3η monomer. A small difference can be seen for residues 166 to 233 when binding the phosphorylated RSRSTpSTP peptide (blue curve) whereas for the monomer binding the same peptide in a dimer simulation with a single peptide (not shown) only residues 204 to 233 showed slightly reduced fluctuations. The second monomer in the dimer simulation showed no differences between binding no peptide, RSRSTSTP, or RSRSTpSTP, respectively. For the 14-3-3η monomer binding either the RLYHSLP or the RLYHpSLP peptides no differences were found when compared to each other or to the monomer without peptides. The 14-3-3η dimer bound to two peptides also showed hardly any difference no matter if both peptides were phosphorylated, unphosphorylated, or mixed.

3.3.2 Principal Component Analysis (PCA)

Next we wanted to find out if the binding of the peptides has an influence on the internal low-frequency modes of the 14-3-3η dimer. For this we performed a PCA analysis on the three concatenated simulations of each system. As baseline we used the main modes of the 14-3-3η dimer without bound peptides. As expected, the first mode is an open-closing movement of the two monomers towards each other without twisting, where the tips of the monomers (at the opposite ends of the dimer interface) move closer to each other when closing and away from each other when opening. The movement of the monomer tips is quite pronounced ($\approx 15\,\text{Å}$ for each tip). Similar first modes were obtained when the dimer is bound at the same time to the RSRSTpSTP and HRYpSTP peptides or to the RSRSTSTP and HRYSTP peptides. All other cases include strong twisting in their first modes.

The second PCA mode is a twisting movement of the two monomers that is coupled to a small opening-closing movement. The opening happens when the monomers twist left, the closing when the monomers twist right. Similar second modes are found in all other dimer simulations no matter which peptides are bound. The third PCA mode is an alternating opening and closing of the two monomers where one monomer is open while the other one is closed. Similar third modes were found when the RSRSTpSTP and HRYpSTP peptides or the RSRSTSTP and HRYSTP peptides simultaneously bind to the dimer, when the dimer binds to a single RSRSTSTP peptide and when the RSRSTSTP and HRYpSTP peptides simultaneously bind to the dimer, though in these cases some twisting movement can be seen as well. The third PCA modes of all other simulations differ from this one and likewise from each other.

Overall the principal components of the 14-3-3η dimer with both peptides either both phosphorylated or both unphosphosporylated at the same time resemble most strongly the ones of the dimer without peptides. The main principal components for dimers bound to a single peptide or when bound to two peptides with differing phosphorylation state differ strongly from the other cases. This shows that the binding of peptides has a strong influence on the dynamics of the full dimer.

3.3.3 Alchemical free energy simulations

The results of the unbiased MD simulations suggest that the phosphorylated peptides are much more tightly coordinated in the 14-3-3 binding groove than in the unphosphorylated peptides. Additionally, the phosphates were better coordinated inside the binding pocket than in bulk water. On the other hand, the side-chain oxygen atoms of the serines in the unphosphorylated peptides were better coordinated in the free state compared to the bound state. Hence, we set out to quantify the binding free energy differences between the phosphorylated and unphosphorylated forms of three peptides bound to the 14-3-3η monomer with the alchemical transformation method. We analyzed three transformations of the RSRSTSTP to the RSRSTpSTP peptide, of the RSRSTSTPNV to the RSRSTpSTPNV peptide, and of the

	exp. ΔG_{bind}		$\Delta\Delta G_{\text{bind}}$		
peptide	S	pS	exp.	alch. sim. (SP1)	alch. sim. (SP2)
RSRSTpSTP	> -24.55†	-35.92†	< -11.37	-14.84 ± 5.27	-41.32 ± 6.42
RSRSTpSTPNV					-39.11 ± 4.14
RLYHpSLP		-42.39‡			-50.80 ± 7.32

TABLE 3.5: Binding free energy differences in kJ mol^{-1} between phosphorylated/unphosphorylated peptides binding to the 14-3-3η domain obtained by alchemical transformations in comparison with experiment at 298.15 K. Values labeled by †were calculated from experimental K_{I} values and those by ‡ were calculated from experimental K_{D} values by Yaffe et al. [104]. The table was taken from ref. [32].

RLYHSLP to the RLYHpSLP peptide, respectively. All transformations were also performed in the backward direction in order to obtain improved convergence.

At physiological pH = 7.2, the phosphate group exists predominantly in the dibasic form (–OPO$_3^{-2}$, further called SP2) and not in the monobasic form (–OHPO$_2^{-2}$, further called SP1) (phosphoserine p$K_{\text{a}} = 5.6$[356]). Therefore, we used the dibasic form in most of our simulations but tested also the monobasic form for the RSRSTSTP to RSRSTpSTP transformation.

The results are listed in table 3.5. The experimental value for the absolute binding free energy of the unphosphorylated RSRSTSTP peptide represents a lower margin to the real binding free energy, since the value could not be precisely determined in the experimental assay [104]. Therefore it is likely much less negative and the binding is therefore much weaker than the binding to the phosphorylated RSRSTpSTP peptide. To our knowledge there exist no experimental binding free energies or binding constants for unphosphorylated motif I, II and III peptides [29, 338]. The only available binding data for unphosphorylated peptides binding to 14-3-3 proteins exist for a range of peptides whose sequences are unrelated to the known 14-3-3 binding motifs [103, 110]. Thus, one can expect that the binding free energy for the unphosphorylated peptides studied here is close to zero. The lower experimental limit of the equilibrium binding constant is 50 µmol (for RSRSTSTP binding 14-3-3η) [104] resulting in a minimum binding free energy of -24.55 kJ mol^{-1}.

All binding free energy differences reported in this study are defined as the binding free energy of the phosphorylated peptide minus the one of the unphosphorylated peptide $\Delta\Delta G_{\text{bind}} = \Delta G_{\text{bind}}^{\text{phosphorylated}} - \Delta G_{\text{bind}}^{\text{unphosphorylated}}$. We will discuss the implications of the reported values in the discussion section.

3.4 Discussion

The aim of this work was to find out mechanistically why the phosphorylated forms of 14-3-3 binding peptides bind stronger to 14-3-3 domains than their unphosphorylated counterparts. We selected several systems for which thermodynamic and structural data are available. Thus, we simulated the 14-3-3η monomer with the RSRSTSTP/RSRSTpSTP and RLHYSLP/RLYHpSLP peptides. Additionally we studied the 14-3-3η dimer binding a single RSRSTSTP/RSRSTpSTP peptide and the dimer binding two peptides at the same time (RSRSTSTP + HRYSTP, RSRSTSTP + HRYpSTP, RSRSTpSTP + HRYSTP, RSRSTpSTP + HRYpSTP). Our results suggest that there exist (if at all) only small differences between a single peptide binding to a 14-3-3η

monomer or to a dimer in terms of the coordination of the phosphate oxygens, the hydrogen bonds, and the contacts between the peptide and the protein, no matter whether the peptide is phosphorylated or not. However, the simulations provide a clear picture why the complex between the 14-3-3η domain (no matter if as monomer or dimer) and the phosphorylated RSRSTpSTP peptide is much more favorable than with the unphosphorylated RSRSTSTP peptide. The main reasons are the strong interactions of the phosphate group with charged residues of the 14-3-3η domain, the strengthening of additional hydrogen bonds due to close peptide-protein contacts of the phosphorylated RSRSTpSTP peptide, and the better coordination of the phosphoserine side chain, as well as the weaker coordination of the serine side chain in the bound state compared to the free state. Similar arguments apply to the case when either a phosphorylated RLYHpSLP peptide or an unphosphorylated RLYHSLP peptide (motif 2) binds to a 14-3-3η monomer. We also observed contacts between the arginines at positions -3, -4 and -5 next to the phosphoserine and the phosphate group that likely enhance its coordination in the bound state. This would explain why the most common binding motifs almost always contain arginine residues at positions -3 (motif 1) and -4 (motif 2) [104–106]. The same effect will likely also exist for motif 3 peptides that also usually contain an arginine at position -3 from the phosphoserine/phosphothreonine [105].

We performed simulations of all systems three times over 1 µs each. As can be seen from the hydrogen-bonding patterns (compare Fig. 3.3, A.3, A.4, A.5, A.6) there exist certain differences between repeated simulations. Even hydrogen bonds that include atoms from the phosphoserine residue vary to some extent. Overall this suggests that 1 µs long simulations are too short to sample all important dynamics of the systems. By integrating the results from three independent simulations, we suggest that our findings are an approximation to the real system properties. This is even more relevant for the simulations of the 14-3-3η dimers, since for these many more interactions and also additional dimer movements play a role in the system dynamics.

The binding free energy differences computed using the alchemical free energy transformation method are more negative than the experimental binding free energies for the phosphorylated peptides. Even if the unphosphorylated peptides had absolute binding free energies of zero (so no binding at all, which could be expected since these were too small to be measured experimentally [104, 338, 339]), the binding free energy difference could not be more negative than the absolute values for the phosphorylated peptides. If the binding free energies of the unphosphorylated peptides are indeed close to zero, the absolute free energies of the phosphorylated peptides would lie in the error range of the results from the alchemical simulations we report here. Reassuringly, the difference computed for the comparison of RSRSTpSTP/RSRSTSTP and RLYHpSLP/RLYHSLP transformations ($9.48\,\mathrm{kJ\,mol^{-1}}$) compares well to the difference of the experimental absolute free energies of the RSRSTpSTP and RLYHpSLP peptides ($6.47\,\mathrm{kJ\,mol^{-1}}$). Possibly, the interactions with the phosphate residue are slightly overestimated in the CHARMM36m force field [192].

Cooperative effects among the two binding sites of 14-3-3 dimers have been suggested based on experimental observations [104, 339, 341]. We tried to analyze if this cooperativity results from entropic effects of the peptide chain or if it stems from some interactions inside the 14-3-3 dimer. However, no particular differences were observed in the contacts between the peptides and the 14-3-3η monomers in simulations of 14-3-3η dimers with two bound peptides at the same time. When the first bound peptide was the phosphorylated RSRSTpSTP, it did not matter if there was a second peptide and what phosphorylation status the second peptide had. The same result was found for HRYpSTP as the first peptide. Hence, this specific result does not support a model where intra-dimer interactions give rise to cooperativity.

A hint to an explanation of the cooperativity between the binding of two peptides in addition to entropic effects of the peptide could be the following: The results from the principal component analysis of the various 14-3-3η dimer simulations show a substantial influence of the peptide binding on the low frequency modes of the dimer. The binding of a single peptide, no matter if phosphorylated or not, as well as of two peptides with different phosphorylation states alters dimer dynamics compared to what is observed when no peptide is bound. Instead, when two peptides with the same phosphorylation state are bound then the normal dimer dynamics were observed. Thus, the internal conformational dynamics of the dimer are left unchanged. This could explain to some extent why it is favorable for the 14-3-3η dimer to bind two phosphorylated peptides at the same time and therefore the cooperativity of the binding.

In summary, we found that studying complexes of 14-3-3 domains with phosphorylated vs. non-phosphorylated peptides by means of molecular dynamics simulations is quite challenging. One µs simulations appear not fully converged in terms of conformational dynamics, since repeated simulations show slightly different binding behavior of the peptides. Alchemical free energy calculations involving doubly charged groups such as phosphate remain challenging even if one applies tricks such as including a mirror peptide where the perturbation is performed in the opposite direction or corrections for electrostatic effects. Nonetheless, a clear picture emerged from our simulations whereby the 14-3-3 binding groove provides a unique superior coordination for phosphorylated peptides compared to their non-modified counterparts highlighting biomolecular recognition.

3.5 Acknowledgement

This project was supported by Deutsche Forschungsgemeinschaft grant He3875/14-1.

Molecular graphics and analyses were performed with UCSF Chimera [343], developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

Special thanks go to V. Gapsys for providing hybrid structures for phosphorylated serines for the CHARMM36m force field to be used in the PMX [267, 346] program.

3.6 Supporting Information

The supporting information is shown in the appendix of this thesis in chapter A.

3.7 Addendum

3.7.1 Additional analysis which did not end up in the manuscript - Residue couplings inside and between monomers

Experimentally, it has been observed that binding of a long peptide to both binding sites of a 14-3-3 dimer at the same time has a stronger binding affinity than binding to a single site [104, 339–341]. We posed the question whether there exists conformational coupling between the two monomer units and how this is affected by the phosphorylation status of the peptide. The residue correlations were calculated using the "corr" function of the gRINN tool [357]. Based on the interaction energies between all residues the tool first calculates the interaction energy correlation between two pairs of residues using only interactions which have a strength of 1 kcal mol^{-1} or stronger. Additionally, a correlation cutoff of 0.4 was used. This new information is then used to calculate the residue correlation which is a sum of all interaction energy correlations which include the specific residue. By way of design, the residue correlation value can be larger than 1 (since it is a sum of multiple interaction energy correlations each having a value in the range 0 to 1). As input all non-covalent interaction energies (interaction energies between non-neighboring residues) were provided. The whole interaction energy analysis was done similar to Kong and Karplus, 2009 [358]. We tested different numbers of simulations frames for the analysis and in the end we used every 20th frame which showed a converged result in all cases. We analyzed each of the three repeats of every simulation on its own and took the mean over all correlations afterwards.

The 14-3-3η monomer binding the phosphorylated RSRSTpSTP peptide shows an overall higher total correlation between the monomer residues (199.96) as the monomer binding the unphosphorylated RSRSTSTP peptide (173.32) and more residues are correlated as can be seen in the upper plots of Fig. 3.5. We performed the same analysis with a monomer without a bound peptide and did not obtain any correlations. The residue LYS125 shows the strongest correlations in the case of a monomer binding the RSRSTSTP peptide, these are much weaker when binding a phosphorylated peptide. This residue mostly binds the side chain of the threonine at position +1 of the serine/phosphoserine. Another residue which has strong correlations in both cases is LYS50. It binds the negative oxygen atoms of the peptides C-terminal (here the proline at position +2 of the serine/phosphoserine in the peptide.) which could thus be an artifact due to a short peptide and would maybe not be this present when simulating the system with longer peptides. This could also be an additional reason why hardly any unbinding of the unphosphorylated peptides in the simulations is observed.

The corresponding clustered graphs (Fig. 3.5 bottom) clearly display one big cluster of residues (which are all interconnected) for the unphosphorylated and the phosphorylated case as well. These clusters are of similar size (phosphorylated: 32 residues, unphosphorylated: 27 residues) and contain ten residues which are present in both clusters. Eight of them are binding pocket residues (LYS50, ARG57, LYS125, ARG132, GLU136, LEU177, ASN178, GLU185) and the other two (PRO36, ASP41) lie on the connection of helices two and three of the 14-3-3η monomer. Both clusters also include two more binding interface residues each which differ in both cases (phosphorylated: ARG61, TYR130; unphosphorylated: ASP129, TYR133). The cluster in the phosphorylated case also includes a few residues (ARG4, ARG12, GLU15, GLU18) which sit in the first helix of the 14-3-3η monomer and are mostly interacting with residues sitting in the helices two and three (ALA24, GLU32, LEU37,

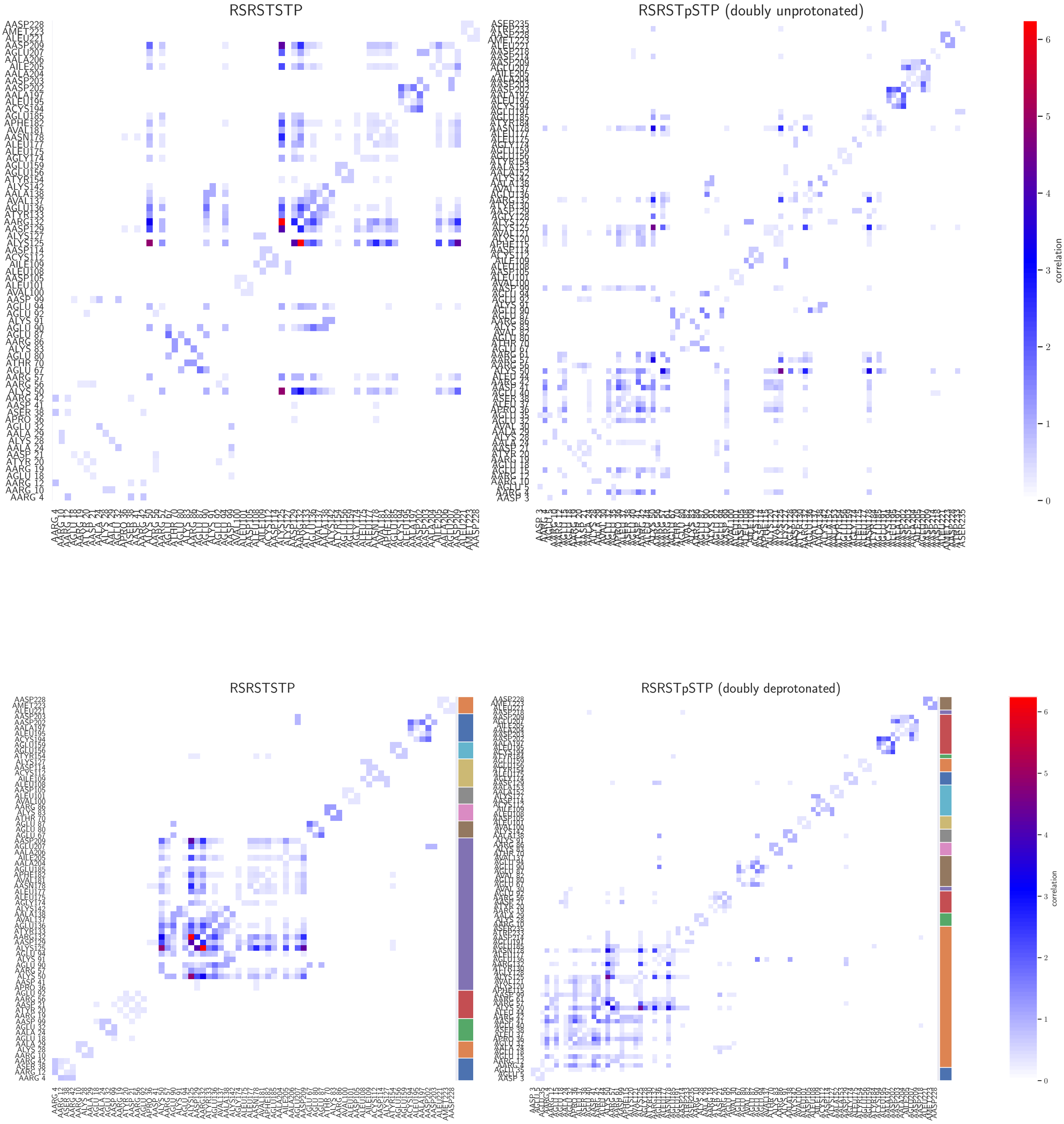


FIGURE 3.5: Top: Residue correlation in the 14-3-3η monomer when bound to the RSRSTSTP/RSRSTpSTP peptides. Bottom: Clustered residue correlation in the 14-3-3η monomer when bound to the RSRSTSTP/RSRSTpSTP peptides using markov clustering.

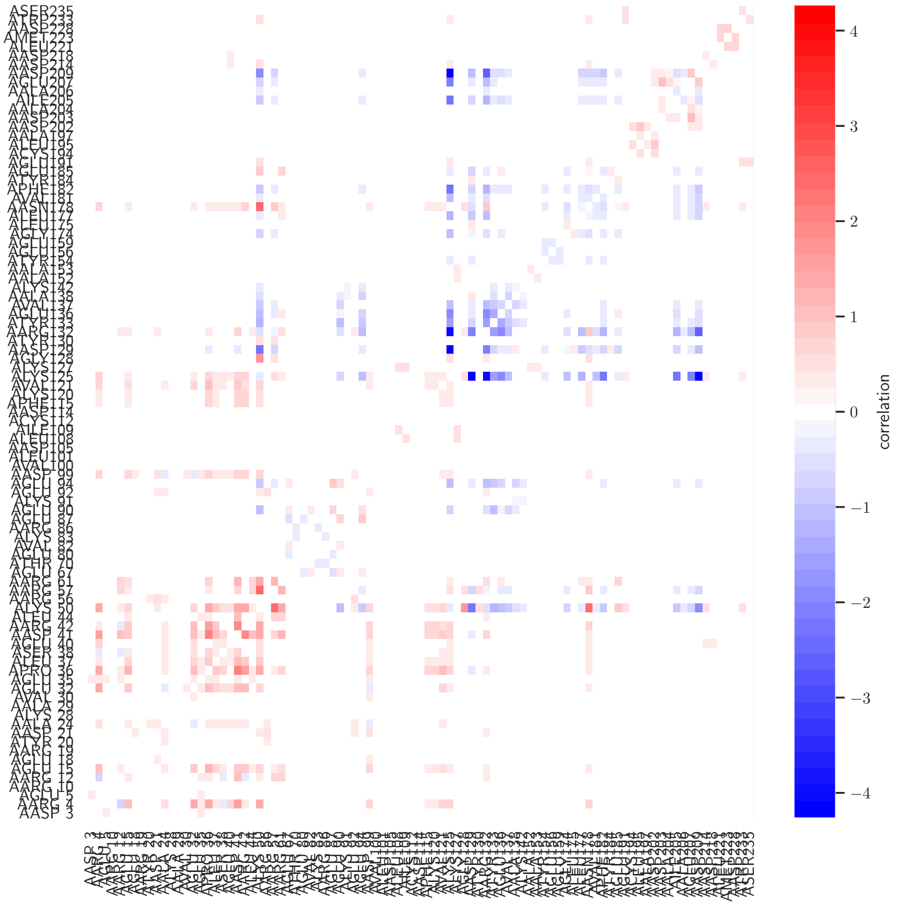


FIGURE 3.6: Difference between residue correlations in the 14-3-3η monomer with RSRSTSTP and RSRSpSTP peptides. Red/positive values mean that the correlations are higher in the simulations with the phosphorylated RSRSTpSTP and blue/negative values that the correlations are higher in the simulations with the unphosphorylated RSRSTSTP peptide.

GLU40, LEU44 and PRO36 and ASP41 which have been named before). Thus the interconnection between the helices one and two with the binding pocket is stronger than in the case of a bound unphosphorylated peptide. All other differences between the clusters are residues far away from the dimer interface. A few clusters are completely the same in both cases. These are mostly clusters of residues which are either inside the same helix or coupling two helices to each other. So they are relevant for the stability of the tertiary structure of the 14-3-3η monomer. One of the clusters could be interesting for a further analysis. It contains the same residues in both cases (ARG19, TYR20, ASP21, ARG56, GLU92). These are all dimer interface residues. In the case of a 14-3-3η monomer binding the phosphorylated RSRSTpSTP peptide the residues in this cluster show some correlations with residues in the big cluster described before which is not the case when binding the unphosphorylated RSRSTSTP peptide. This, together with the stronger interaction of helices one and two with the binding pocket, could be a hint of a relevant connection between the binding pocket and the dimer interface (which includes many residues of helices one and two) when binding phosphorylated peptides which does not exist, at least this strong, when binding unphosphorylated peptides. This has to be investigated further.

We also analyzed the residue correlations of a 14-3-3η dimer binding an unphosphorylated RSRSTSTP or a phosphorylated RSRSTpSTP peptide. Here the dimer with the RSRSTSTP peptide (329.62) has an overall higher total correlation than the dimer binding the RSRSTpSTP peptide (230.90) which is the opposite of the results we obtained for the monomers binding the peptides. The same holds true when we compare the total correlation only inside the monomer which actually binds the peptide in these dimer simulations (with RSRSTSTP peptide: 110.69, with RSRSTpSTP peptide: 83.10), which is further called monomer A. Comparing the two cases further we see that there exist only weak correlations that include binding pocket residues when binding a unphosphorylated RSRSTSTP peptide compared to when binding a phosphorylated RSRSTpSTP peptide.

The residue correlations which we obtained in simulations of a dimer with a single RSRSTSTP peptide are very different to the ones we obtained for a monomer with the RSRSTSTP peptide. While LYS125 was the strongest correlated residue in the latter, this residue has nearly no correlations in the former. Additionally, for the dimer GLU90 has the strongest correlations while it has very few weak correlations for the monomer. These results would suggest that there exist no single specific binding mode for the unphosphorylated RSRSTSTP peptide with 14-3-3η.

While for simulations where an unphosphorylated RSRSTSTP peptide was bound the monomer and dimer simulations lead to totally different residue correlations, simulations of a 14-3-3η dimer binding a single phosphorylated RSRSTpSTP peptide are very similar to simulations of a 14-3-3η monomer binding the same peptide even though the correlations inside the monomer which binds the peptide (further called monomer A) are overall less and weaker. The opposite is the case for simulations of a 14-3-3η dimer binding two phosphorylated peptides (RSRSTpSTP and HRYpSTP) at the same time. Here the correlations inside the monomer binding the RSRSTpSTP peptide are similar but much stronger than in the case of a monomer binding the RSRSTpSTP peptide and also new residue correlations are found which do not exist in the latter case. Interestingly strong correlations between ASP129 and other binding pocket residues (e.g. LYS50, ARG57 and LYS125) are found which are not present when only a single phosphorylated RSRSTpSTP peptide is bound no matter if to a dimer or a monomer.

In both cases where a dimer with a single peptide was simulated correlations between both monomers are found. For the dimer binding a single phosphorylated RSRSTpSTP peptide these are most strongly between the binding pocket residues of the monomer binding the peptide (further called monomer A) and residues in the first and second helix of the other monomer (further called monomer B) but these are no dimer interface residues. Another cluster of correlations exists between dimer interface residues of monomer A and B and further residues in the second last helix of monomer B. A similar cluster is found for the dimer binding the RSRSTSTP peptide which is also the biggest cluster in that case. The binding pocket residues of monomer A are mostly found in one cluster in the case of a dimer binding RSRSTpSTP which also includes residues of the helices one and two and of the connection between helices two and three of monomer B. Even more binding pocket residues of monomer A in a single cluster are found for the dimer binding RSRSTSTP but their correlations are much weaker than in the former case and there are only very weak correlations with residues of monomer B (which are not even in the same cluster).

For the dimer binding the two phosphorylated peptides RSRSTpSTP and HRYpSTP the picture is different when compared to the dimer binding a single RSRSTpSTP peptide. In the former case there exists a cluster of correlations (the main cluster here) between the binding pocket of the monomer binding the RSRSTpSTP peptide (further called monomer A), the dimer interface of monomer A and the binding pocket of the monomer binding the HRYpSTP peptide (further called monomer B) even though the correlations inside the binding pocket of monomer A are much stronger than the ones inside the binding pocket of monomer B or between the binding pockets or with the dimer interface of monomer A. There exists another cluster with residues in the dimer interface of monomer B which is not completely isolated but has correlations with the binding pocket residues of monomer B.

When studying the residue correlations for simulations of a 14-3-3η dimer binding the unphosphorylated RSRSTSTP (bound at monomer A) and HRYSTP (bound at monomer B) peptides at the same time one huge cluster is found which includes binding pocket residues of monomer A and B and dimer interface residues as well. The correlations inside the binding pockets are much weaker than when binding the two phosphorylated peptides. The correlations between the two binding pockets are similar in strength. The correlations of the binding pockets with the binding interfaces are weak and there is no evidence for a stronger interaction with one side of the dimer interface than with the other

These findings suggest that the cooperativity between the binding pockets which is found in experiment when binding two phosphorylated peptides [104, 339, 341] is maybe provoked by an information transfer from the binding pocket of monomer A over the dimer interface at helices one and two of monomer A to the binding pocket of monomer B but not over the dimer interface at helices one and two of monomer B. This would also fit into the picture of a gatekeeping mechanism which was found for this system, where the RSRSTpSTP peptide is the gatekeeper for the HRYpSTP peptide due to which the strength of the interaction between the HRYpSTP peptide and the 14-3-3 dimer increases [340]. The mechanism seems to be weak as the strongest correlations were found inside each binding pocket. When binding two unphosphorylated peptides there was no evidence found for such a specific information transfer. This could be a hint that specifically the binding of phosphates is relevant to start this.

Overall it is clear that simulations of a 14-3-3η monomer binding peptides are no good substitutes for simulations of 14-3-3η dimers binding the same peptides even though this would greatly reduce the needed computation time.

This part did not end up in the manuscript because we realized deviations between taking the mean of correlations after analyzing the three simulations on their own and directly analyzing a combined trajectory. The reason is likely the cutoff of $1\,\mathrm{kcal\,mol^{-1}}$ when calculating the interaction energy correlation. When analyzing a singly trajectory it can occur that other residue pairs fall into the cutoff as when analyzing the combined trajectory. And since the results were not very conclusive, we did not include them in the manuscript.

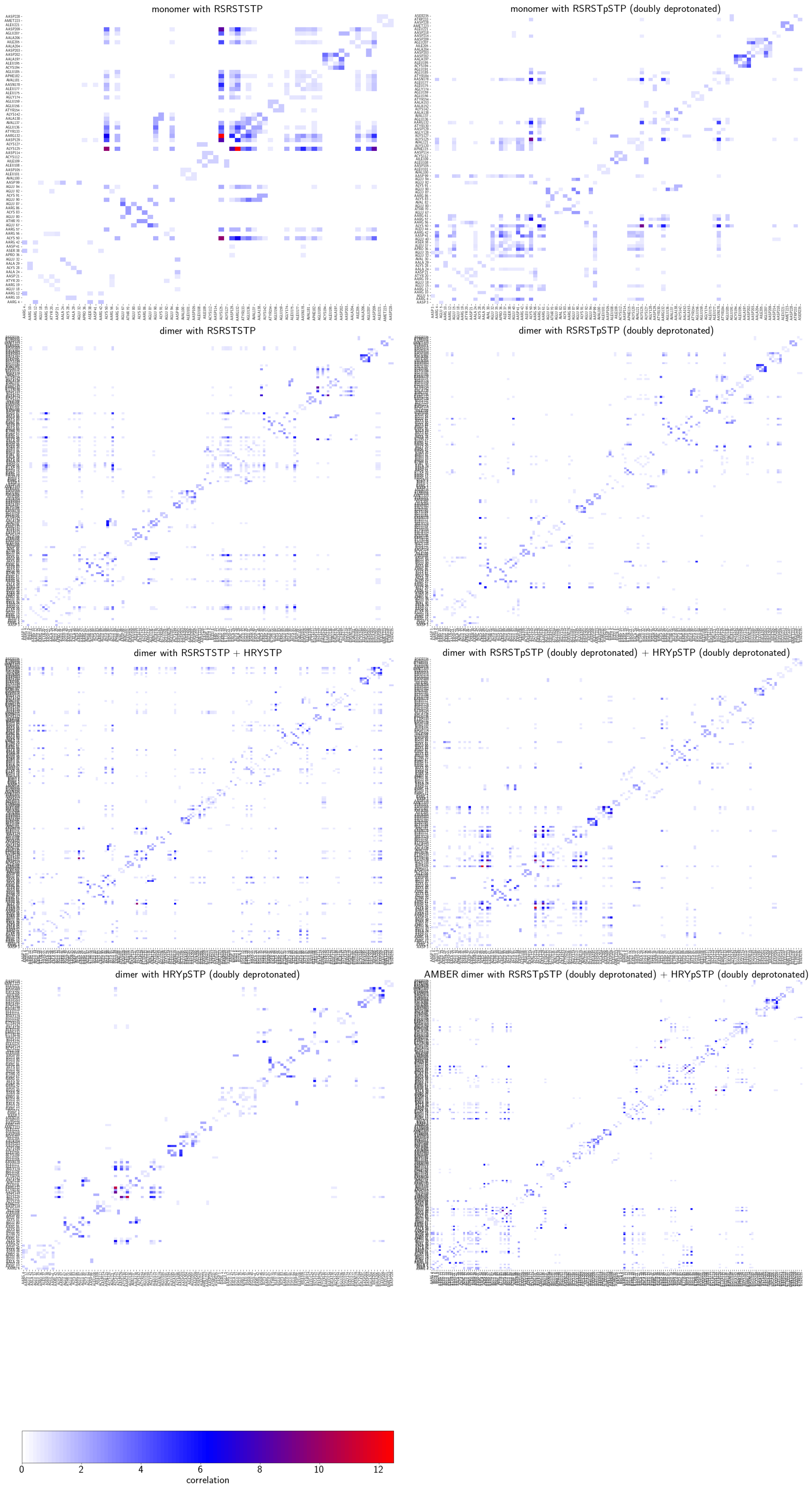


FIGURE 3.7: Residue correlation in the 14-3-3η monomer and dimer when bound to one or two peptides.

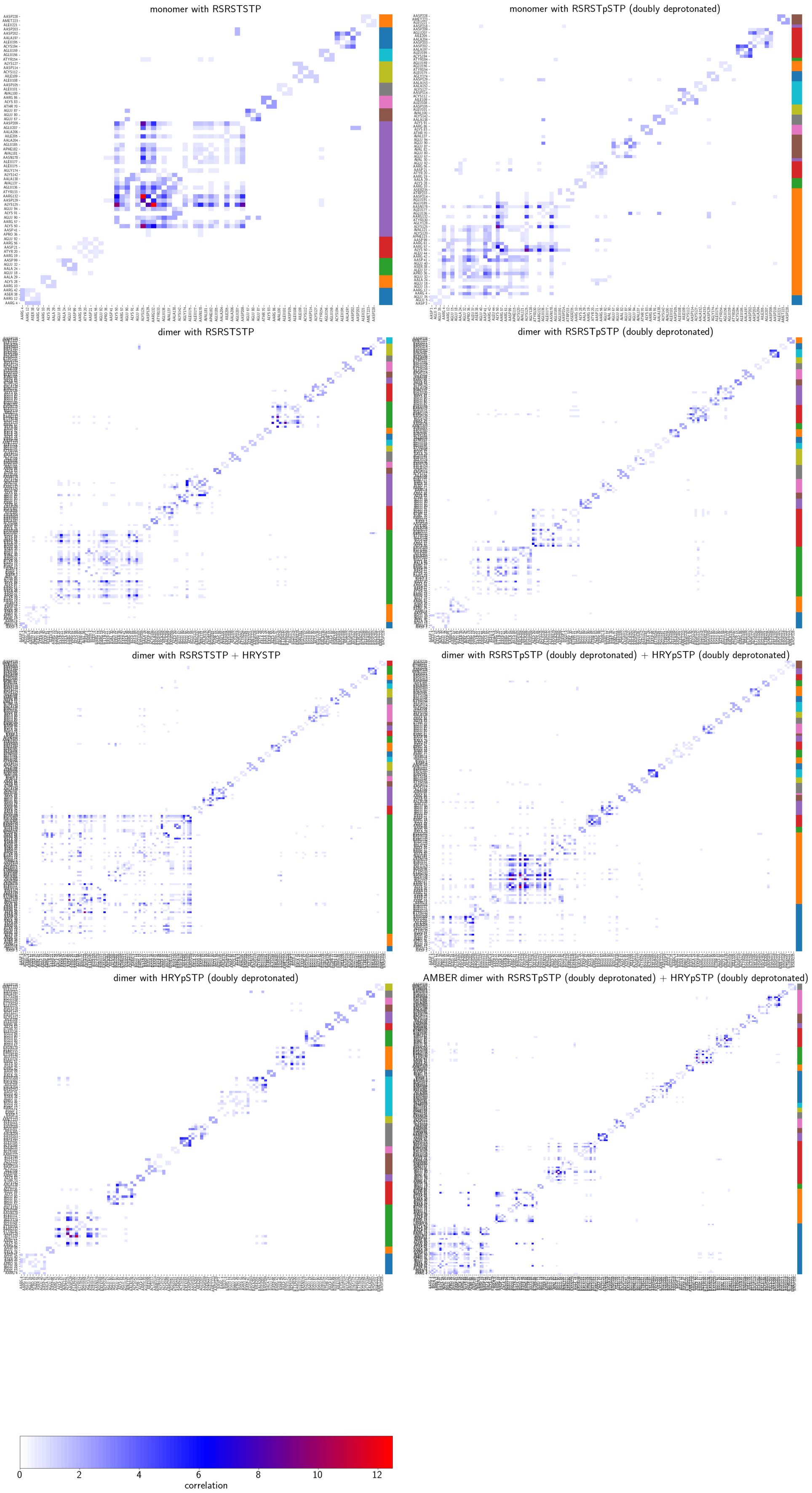


FIGURE 3.8: Clustered residue correlation in the 14-3-3η dimer when bound to one or two peptides using Markov clustering.

Chapter 4

How peptides bind to PDZ domains

This chapter introduces our approaches on understanding the interaction of PDZ domains with phosphorylated and unphosphorylated peptides and what influences specific residues in the PDZ domains have on these interactions. My contributions were the choice of the systems to study and the simulations to perform, performing the simulations, analyzing the data and writing the first draft of the manuscript. In the following sections the manuscript is presented. The abstract is given in the introduction chapter 1.

4.1 Introduction

PDZ (PSD-95/Discs-large/ZO-1) [117] domains are highly abundant protein modules involved in many protein-protein interactions and play an important role in the assembly of supramolecular complexes and signal transduction [120, 122–126]. PDZ domains usually bind to the C-terminal ends of their partner proteins via the three main sequence motifs X[S/T]XΦ-COOH (motif 1), XΦXΦ-COOH (motif 2) and XEDXΦ-COOH (motif 3) [29] where X means any residue and Φ means a hydrophobic residue. The interactions are often controlled by phosphorylation of the peptide at the serine or threonine residue of motif 1.

Previous *in silico* studies have already adressed the binding of peptides to PDZ domains. Blöchliger et al. studied the binding of the Acetyl-EQVSAV peptide to the second PDZ domain of protein tyrosine phosphatase 1E [25]. Their results showed that non-native salt bridges helped in kinetically stabilizing the encounter complex during the binding and they suggested that binding of charged peptides to PDZ domain can be steered by non-native interactions. The allosteric effects of the same interaction on the structure of the PDZ domain were studied by Morra et al. [359]. They found that in PDZ2 a dynamic and energetic reorganization takes place when a ligand binds. Panel et al. studied the binding of C-terminal peptides to the TIAM1 PDZ domain by calculating the relative binding free energy differences using alchemical simulations and additive and polarizable force fields [360] and found that alchemical simulations of C-terminal peptides bindign to the TIAM1 PDZ2 domain can reproduce experimental results within an error range between $1 - 3\,\mathrm{kcal\,mol^{-1}} \approx 4 - 13\,\mathrm{kJ\,mol^{-1}}$ when using additive force fields. Harish et al. designed and converted a low-affinity tetrapeptide to a tight binding one and studied its interactions with the PSMD9 PDZ domain by docking, molecular dynamics simulations and MMPBSA calculations [361]. They found a so far unknown occupancy for cysteine at the P_{-2} position that increases the affinity of the peptide to the PDZ domain.

Here, in contrast to Blöchliger et al. we were not only interested in the binding of the EQVSAV peptide to the PDZ domain but how the binding of a PDZ domain to the C-terminus of a protein is affected by mutation of the central residue or by the phosphorylation of this terminal peptide. For this reason we studied the binding of a six amino acid long peptide derived from the C-terminus of the rap guanine nucleotide exchange factor 6 (RAPGEF6, other names: PDZ-GEF2, RA-GEF2) to the second PDZ domain of human tyrosine-protein phosphatase non-receptor type 13 (PTPN13, other names: hPTP1E, PTP-BAS, PTPL1), which is the same system that was studied by Blöchliger et al. [25]. The peptide has the sequence $\text{EQVSAV}_{\text{COO}^-}$ and can be phosphorylated at the serine at position -2 from the C terminus. Our plain molecular dynamics simulations showed that the of the $\text{EQVEAV}_{\text{COO}^-}$ peptide and the phosphate group of the phosphorylated $\text{EQVpSAV}_{\text{COO}^-}$ peptide bind strongly to ARG79 of hPTP1E. Thus, we additionally studied the influence of this residue on the binding of the unphosphorylated and phosphorylated peptides to the hPTP1E PDZ2 domain. ARG79 was also shown to play an important role for peptide binding by Blöchliger et al. [25]. To test this in simulations, we mutated the arginine to alanine and re-performed our simulations in order to study the importance of this specific residue interaction.

We also wanted to study if the $\text{EQVSAV}_{\text{COO}^-}$ peptide binds in a similar or different manner to other PDZ domains, especially if they have no arginine at the position similar to ARG79 in hPTP1E PDZ2. Using the BioGRID interaction database v3.5 [163] and the STRING database [164] we found the first PDZ domain of the membrane-associated guanylate kinase, WW and PDZ domain-containing protein 1 (MAGI1, other names: AIP3, BAP1, WWP3) and used it as our test case. This protein was shown to bind RAPGEF6 as well as the rap guanine nucleotide exchange factor 2 (RAPGEF2, other names: PDZ-GEF1, CNrasGEF, nRapGEP, RA-GEF1) which also contains a C terminus consisting of the sequence $\text{EQVSAV}_{\text{COO}^-}$. Furthermore, the MAGI1 PDZ1 domain does not contain a lysine at the position equivalent to LYS72 in hPTP1E PDZ2 that may also form multiple interactions with glutamate and phosphoserine. By introducing the mutation Q85R in MAGI1 PDZ1 (position 85 in MAGI1 PDZ1 is equivalent to position 79 in hPTP1E PDZ2), we then studied all four possible combinations of arginine and lysine being present or absent at these specific positions.

We present results from a comprehensive series of plain MD simulations for complexes of PDZ domains bound to the peptides $\text{EQVSAV}_{\text{COO}^-}$, $\text{EQVEAV}_{\text{COO}^-}$ and $\text{EQVpSAV}_{\text{COO}^-}$. Also, we performed alchemical transformation simulations to obtain binding free energy differences between the phosphorylated and unphosphorylated peptides as well as between the peptide with glutamate and the one with serine. Additionally, we performed two parallel cascade selection molecular dynamics simulations to calculate the absolute binding free energy differences for the $\text{EQVSAV}_{\text{COO}^-}$ and $\text{EQVpSAV}_{\text{COO}^-}$ peptides binding to the hPTP1E PDZ2 domain.

4.2 Materials and Methods

4.2.1 Structure preparation

When we started this project there was no crystal structure available for the MAGI1 PDZ1 domain binding the RA-GEF2 peptide. Thus another structure (PDB ID: 2KPL [362]) was used, where the MAGI1 PDZ1 domain binds a peptide from human papillomavirus (HPV) E6. The first model from the set of NMR structures in the PDB-file was used as the starting conformation for the molecular dynamics (MD) simulations. Afterwards the residues of the HPV E6 peptide were exchanged to the ones of the RA-GEF2 peptide (EQVSAV$_{\mathrm{COO}^-}$, in the following we will skip the COO$^-$ for ease of reading). Both the MAGI1 PDZ1 NMR structure (PDB ID: 2KPL [362]) as well as the X-ray diffraction structure for the second PDZ domain from human PTP1E (hPTP1E) in complex with a RA-GEF2 peptide (PDB ID: 3LNY [363]) were taken from the Protein Data Bank (PDB) (www.rcsb.org) [6, 7].

The protonation states of histidine residues were calculated using PROPKA3.0 [364, 365] as part of the PDB2PQR [366] webserver. They are different to the ones in the crystal structure of 3LNY since the structure was determined at pH = 6.8, while the experimental values [156] for the binding free energy were measured at pH = 7.2 which we wanted to reproduce in our simulations.

To perform alchemical free energy simulations for the change of phosphorylated to unphosphorylated peptides or the opposite, we needed topologies and hybrid structures that include the unphosphorylated and phosphorylated residue in a single file (similar to our previous project [32]). Such files were kindly given to us by V. Gapsys for the CHARMM36m force field [192] for the use in the PMX program [267, 346]. This program incorporates a hybrid structure library and allows to create topologies with hybrid structures.

Since we found that arginine located at position 79 of the hPTP1E PDZ2 domain could have a strong influence on the binding of the phosphorylated peptide, we replaced this residue by alanine using the rotamers tool in UCSF Chimera [343] for additional simulations and created a hPTP1E PDZ2 R79A structure. We also exchanged glutamine at position 85 to arginine of the MAGI1 PDZ1 domain to obtain a MAGI1 PDZ1 Q85R structure. For the simulations with the EQVEAV peptides we exchanged the serine in the RA-GEF2 peptide EQVSAV to glutamic acid using the rotamers tool of UCSF Chimera [343]. For the simulations with the EQVpSAV peptide we added the phosphate group to the serine side chain using the build structure tool of UCSF Chimera [343].

4.2.2 Molecular dynamics (MD) simulation protocol

We followed the same plain MD simulations protocol as we used in our previous study [32] but describe it here for completeness. All simulations in this project were performed using the GROMACS [4] 2018.8 software package. The properties of the protein and peptide atoms were provided by the CHARMM36m [192] force field and we used the CHARMM-force-field-modified TIP3P [224] [190] water model for the water molecules. We generated the hydrogen atoms with GROMACS internal tool pdb2gmx. The simulations were performed using a time step of 2 fs and periodic boundary conditions (PBCs). For the long-range electrostatic interactions the particle-mesh-Ewald method [9] was used and we set the non-bonded interaction cutoff to 12 Å. We used the LINCS [347] algorithm to constrain all bonds for our alchemical simulations while for plain MD and PaCS-MD we constrained only the bonds that involved hydrogen atoms. Additionally, we applied long range dispersion corrections for energy and pressure.

For the plain MD simulations a triclinic box was created such that the minimal distance of every atom of the solute is at least 1.5 nm from the box edges resulting in box dimensions of around 6.5 nm per side for the hPTP1E PDZ2 and the MAGI1 PDZ1 domain.

For the alchemical simulations with the "double system in a single box setup" [267] a triclinic box was created such that the minimal distance of every atom of the solute is at least 2 nm away from the box edges resulting in a box size of around 13 nm x 7.5 nm x 7.5 nm. One side is much longer than the other two since the simulation box contains two systems at the same time which have to be separated far enough in order to make sure that interactions between the two are of negligible magnitude.

For the alchemical simulations where the bound and free states were studied in separate boxes we created triclinic boxes with a size of around 6.9 nm x 6.9 nm x 6.9 nm for the bound and 5.2 nm x 5.2 nm x 5.2 nm for the free state.

For the parallel cascade selection molecular dynamics (PaCS-MD) simulations a triclinic box with a size of around 8.5 nm x 8.5 nm x 14.1 nm was created. The z-direction was set to a larger value because the binding pocket was aligned along this direction in order to follow the dissociation of the protein and peptide mostly in this direction.

For all simulations the system was first minimized for a maximum number of 50000 steps and an initial step size of 0.01 nm using steepest descent minimization. It was stopped when the maximum force reached a convergence value of 5.0 kJmol^{-1}nm^{-1}. Afterwards the system was solvated in explicit water using the TIP3P [224] water model and neutralized and ionized to a concentration of 150 mmol NaCl (if not mentioned otherwise for a specific simulation) as it was also done in the binding experiments of Toto et al. [156] for the binding between the hPTP1E PDZ2 domain and the EQVSAV peptide to which we wanted to compare our results. This was followed by another energy minimization using the steepest descent method for a maximum number of 50000 steps, an initial step size of 0.01 nm and a convergence value of 500.0 kJmol^{-1}nm^{-1} whereby the heavy protein and peptide atoms were kept rigid using position restraints.

Next we thermalized the systems for 500 ps each at 100 K, then at 200 K and finally at 283 K (final temperature as in the experiment [156]) using the stochastic velocity rescaling algorithm by Bussi et al. [239] with two temperature groups, one for the protein and one for the rest, and a coupling time constant of 0.1 ps for both groups. The system was then further equilibrated for another 500 ps in the NPT ensemble. Here, a Berendsen barostat [238] with a time constant $\tau_p = 2$ ps, an isothermal compressibility of 4.6×10^{-5} bar^{-1} and reference pressure $p = 1.0$ bar was used. Afterwards the contraints on the positions of the protein atoms were removed in three steps (1000 to 100 to 10 to 0 kJ/mol/nm^2) for 400 ps each, from this point the pressure coupling was changed to a Parinello-Rahman barostat [250, 348] with time constant $\tau_p = 5$ ps and reference pressure $p = 1.0$ bar). In the last step all constraints were released and the system was allowed to equilibrate freely. This creation and equilibration protocol was implemented for all simulations.

Plain MD simulations

Unbiased production runs of 1 µs length were performed for all thirteen plain MD simulations (hPTP1E PDZ2, hPTP1E PDZ2 R79A, MAGI1 PDZ1 and MAGI1 PDZ1 Q85R all with the unphosphorylated EQVSAV peptide, the phosphorylated EQVpSAV peptide (monobasic phosphoserine) and the EQVEAV peptide and hPTP1E PDZ2 with the phosphorylated EQVpSAV peptide with dibasic phosphate). Snapshots of the trajectories were recorded every 10 ps. Since we observed unbinding events for MAGI1 PDZ1 with the EQVEAV peptides and hPTP1E PDZ2 R79A with the EQVEAV peptide and partial unbinding for MAGI1 PDZ1 with the EQVSAV peptide as well as for MAGI1 PDZ1 Q85R with the EQVpSAV peptide we repeated these simulations each once. The second simulation of hPTP1E PDZ2 R79A with the EQVEAV peptide showed partial unbinding wherefore we repeated this once more.

Alchemical free energy simulations

The general procedure of the alchemical free energy simulations was performed similar to our previous work [32]. For the alchemical free energy simulations first an unbiased production run of 20 ns was performed. Snapshots of the trajectories were recorded every 10 ps. From this production we discarded the first 4 ns for further equilibration purposes and initial structures for the alchemical transformation simulations were taken every 16 snapshots (160 ps) from the resulting equilibrium trajectory. Afterwards alchemical transformation simulations were started from each of these 100 starting structures for a simulation time of 1 ns each (if not stated otherwise) where the coupling parameter λ was changed from 0 to 1 during the course of the simulation. Because atoms are created and annihilated in these processes abrupt changes of Lennard-Jones and Coulomb potentials can lead to poor convergence. Therefore, a soft-core potential was used that removes the singularities in the potentials. Every time step we saved how the Hamiltonian changed with respect to the coupling parameter λ. The simulations were performed forward and backward, i. e. a change of λ from 0 to 1 and from 1 to 0. This allowed to obtain a better converged resulting binding free energy differences. We then combined the results using the analysis tool of PMX [267]. With this the free energy difference can be calculated via the Bennet acceptance ratio (BAR) [298], the Jarzynski equality [184], and/or the Crooks Gaussian Intersection [351] method. We obtained all our results by using BAR but to get a better level of convergence we only used results when all of the three estimators provided results within 4 kJ mol^{-1}.

Parallel Cascade Selection Molecular Dynamics (PaCS-MD)

Parallel cascade selection MD (PaCS-MD) was used to obtain absolute binding free energies for the binding of the unphosphorylated EQVSAV peptide as well as for the binding of the phosphorylated EQVpSAV peptide with singly charged (SP1) phosphate to the hPTP1E PDZ2 domain. The simulations were performed similar to [27] and we used the center-of-mass (COM) distance as the reaction coordinate to separate the protein and the peptide during the PaCS-MD simulations.

For the PaCS-MD simulations we kept restraints on the C_α atoms of residues 7-10, 58-60, and 86-69 with a strength of 100 kJ/mol/nm^2 to prevent the protein from rotating and moving around.

33 trials (independent runs) for the unbinding of the unphosphorylated peptide and 30 trials for the unbinding of the phosphorylated peptide of PaCS-MD simulations were performed to obtain a reasonable amount of sampling. The starting structures for these trials were taken every 1 ns from the last 30 ns of an initial plain MD simulation with 100 ns length. For each trial, a 1 ns long MD simulation (cycle 0) was started from these initial structures. 30 replicas were started from the 30 snapshots with the largest inter-COM distance between the protein and the peptide. Each replica was simulated for 100 ps in each cycle and we recorded snapshots of the trajectories every 0.5 ps. The simulations of a trial were stopped when the inter-COM distance between protein and peptide exceeded $d = 6$ nm for all the 30 best snapshots of the previous cycle.

We used pyEMMA [367] to estimate and analyze reversible discrete Markov state models (MSMs) from the PaCS-MD trajectories using the relative three-dimensional center of mass positions (3D-COM) of the peptide with respect to the protein as was shown to be a good choice in [27]. All trajectories from the PaCS-MD trials were used but all snapshots with inter-COM distances larger than 5 nm were discarded from the analysis since complete phase-space sampling in the regions far away from the binding pockets is hard to achieve and the protein and the peptide are already completely separated at 5 nm. This resulted in a total analyzed simulation time of 2.620 µs for hPTP1E PDZ2 with the unphosphorylated EQVSAV peptide and of 3.898 µs for hPTP1E PDZ2 with the phosphorylated EQVpSAV peptide with SP1 phosphate.

In order to find a reasonable number of cluster centers or microstates we estimated unvalidated Markov models for various numbers of cluster centers which were discretized using the Kmeans clustering [368] approach. As a heuristic we used the VAMP-2 score [331] (with cross validation). We estimated the MSM lag time as 20 steps for the PaCS-MD of hPTP1E PDZ2 with the unphosphorylated EQVSAV peptide which is equal to 10 ps to perform this analysis. The lag time had to be adjusted later on. Repeated discretization rounds were performed in order to account for the stochastic nature of the clustering algorithm. A reasonable cluster number turned out to be 3000 clusters where the VAMP-2 score reached a plateau as is shown in fig. 4.4.

The trajectory was then transformed into a discretized trajectory along these microstates and MSMs were built for various lag times and the resulting time scales for the 10 slowest processes were plotted to find the optimal lag time to build the MSMs which is usually chosen to be the lag time where the time scales reach a converged plateau. A reasonable lag time to build our MSMs was found to be 40 ps.

The stationary distribution $\pi(\mathbf{r}_j)$ of each microstate j of the Markov state model based on the relative 3D-COM position of the peptide was mapped into a 1D-PMF by

$$W(d_i) = -\frac{1}{\beta} \ln \sum_{d_i - \delta d/2 \leq d(\mathbf{r}_j) < d_i + \delta d/2} \pi(\mathbf{r}_j) \text{ [27].} \tag{4.1}$$

with the inverse temperature $\beta = 1/(k_\mathrm{B}T)$, the position in 3D of microstate j, $\mathbf{r}_j$, and the bin size δd. The bin size was chosen to be $\delta d = 0.05\,\mathrm{nm}$.

4.2.3 Analysis tools

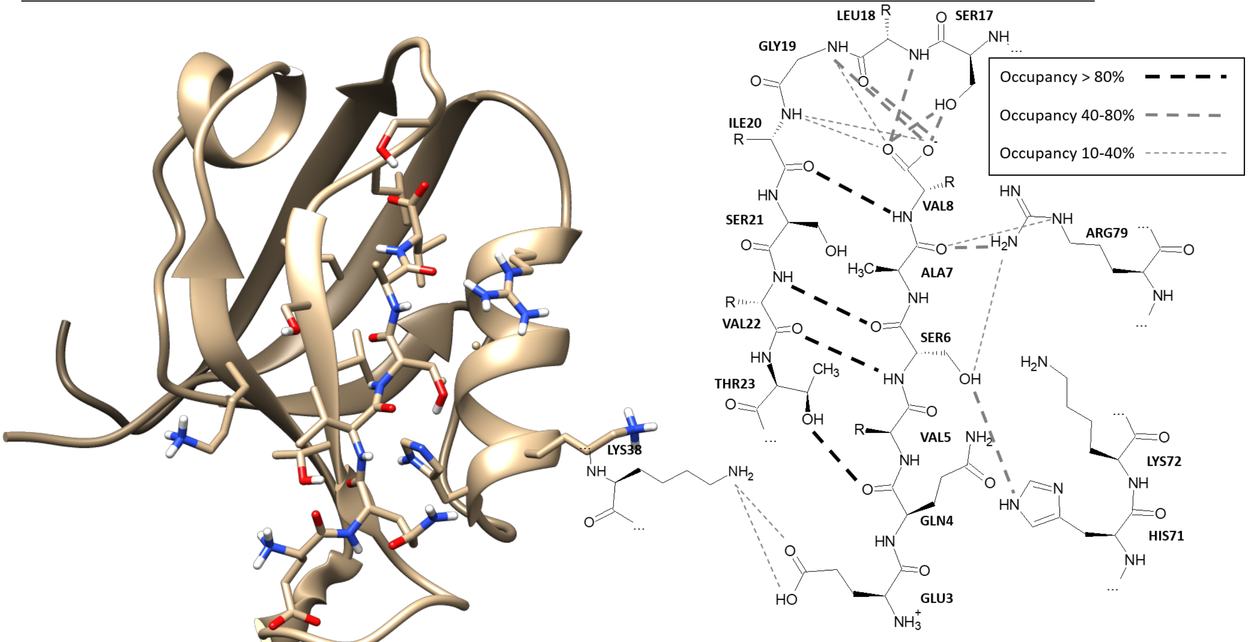
We applied different tools our data in order to analyze the differences between the bound states of hPTP1E PDZ2, hPTP1E PDZ2 R79A, MAGI1 PDZ1 and MAGI1 PDZ1 Q85R in complex with the various peptides. For the analysis of hydrogen bonds and their strength we processed our trajectories with the MDTraj [349] python toolkit and applied the Kabsch-Sander [369] and Baker-Hubbard [350] algorithms.

4.3 Results

In the following peptide residues will be labeled XXX-Pep and protein residues will be labeled XXX-Prot, where XXX is the short acronym of the peptide or protein name. The peptide residue numbering always begins at 3 and ends at 8, since this is the numbering used in the structure PDB ID 3LNY [363].

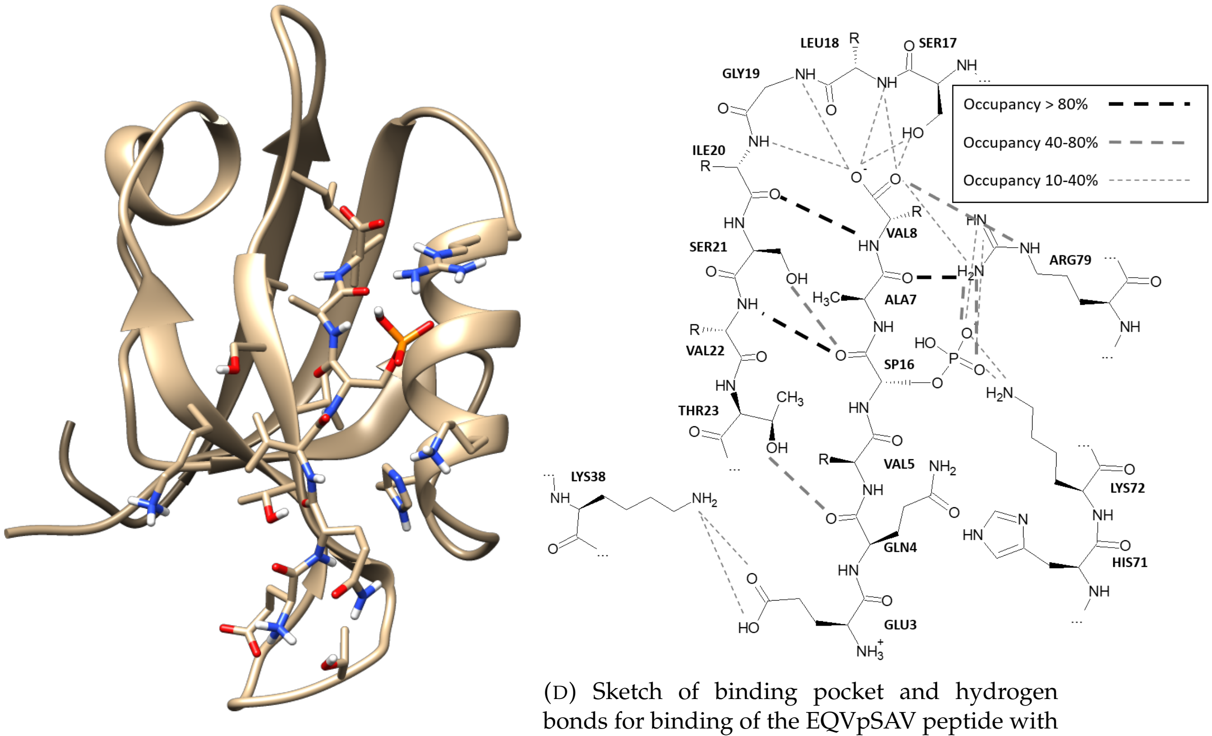
4.3.1 Unbiased molecular dynamics of PDZ domains binding the EQVSAV and EQVEAV peptides

Most of the performed plain MD simulations of the EQVSAV and EQVEAV peptides bound to the PDZ2 and PDZ2 R79A domains of hPTP1E and to the PDZ1 and PDZ1



(A) hPTP1E PDZ2 binding EQVSAV

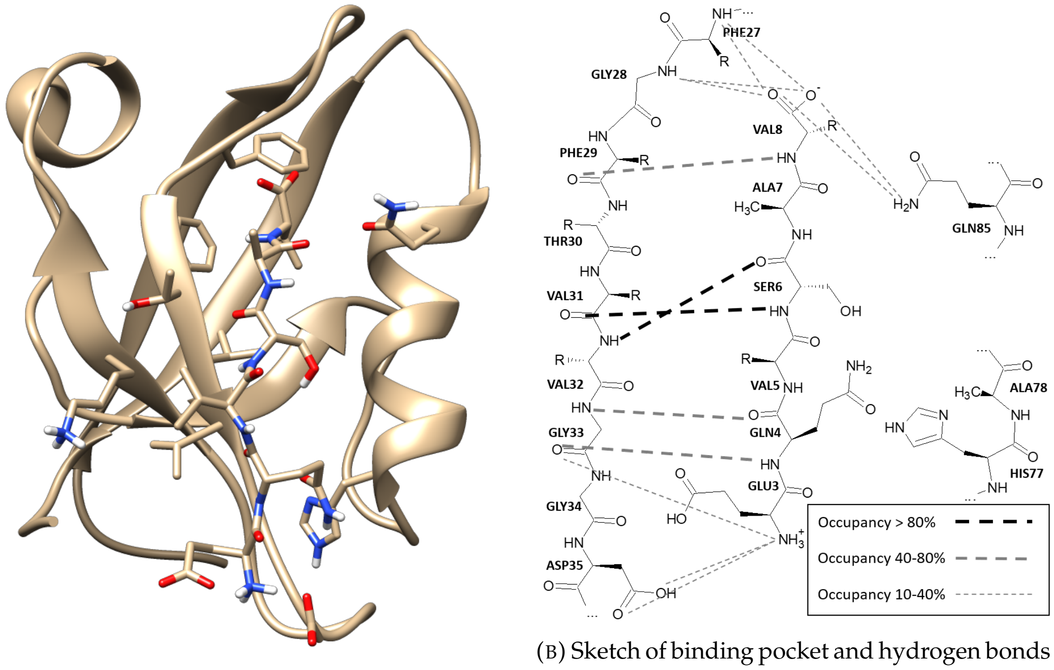
(B) Sketch of binding pocket and hydrogen bonds for binding of the EQVSAV peptide to hPTP1E PDZ2.



(C) hPTP1E PDZ2 binding EQVpSAV with singly charged (SP1) phosphate

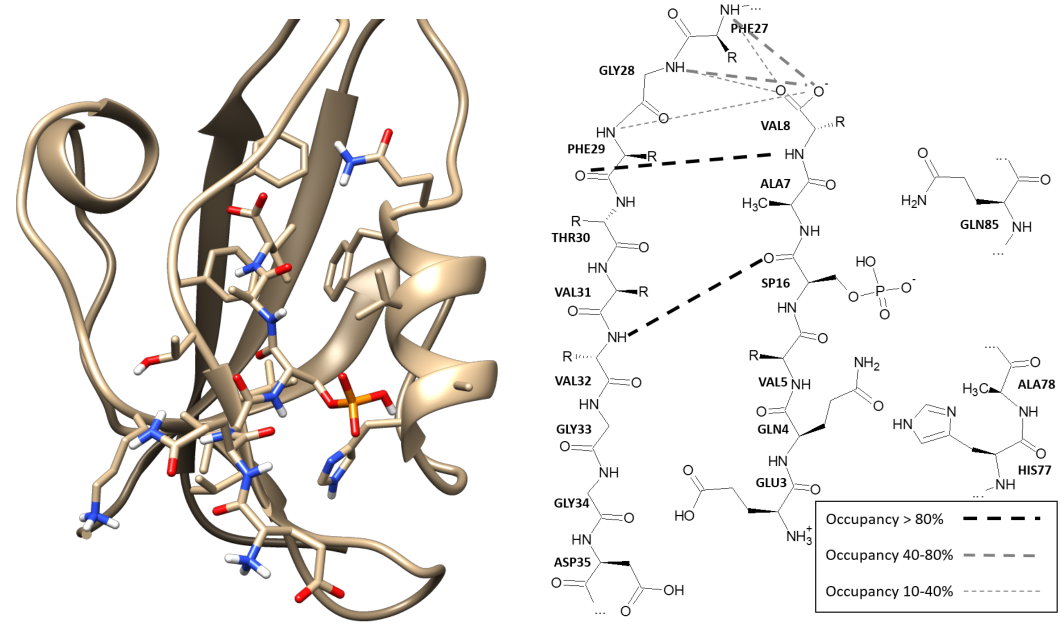
(D) Sketch of binding pocket and hydrogen bonds for binding of the EQVpSAV peptide with singly charged (SP1) phosphate to hPTP1E PDZ2.

FIGURE 4.1: Representative structures of the EQVSAV peptide (top left) and the EQVpSAV peptide with singly charged (SP1) phosphate group (bottom left) binding to the hPTP1E PDZ2 domain. These structures were extracted similar to our previous study [32] (see chapter 3) by finding the centroid, the frame with the highest sum of similarities, of all trajectory frames of 1 µs plain MD simulations with 2D sketches of the binding pocket and the relevant hydrogen bonds between protein and peptide on the right side, respectively. As a distance metric we used the pairwise RMSD and then used the pairwise distances to calculate a pairwise similarity.



(A) MAGI1 PDZ1 binding EQVSAV

(B) Sketch of binding pocket and hydrogen bonds for binding of the EQVSAV to MAGI1 PDZ1.



(C) MAGI1 PDZ1 binding EQVpSAV with singly charged phosphate group

(D) Sketch of binding pocket and hydrogen bonds for binding of the EQVpSAV with singly charged (SP1) phosphate group to MAGI1 PDZ1.

FIGURE 4.2: Representative structures of the EQVSAV peptide (top left) and the EQVpSAV peptide with singly charged (SP1) phosphate group (bottom left) binding to the MAGI1 PDZ1 domain. These structures were extracted similar to our previous study [32] (see chapter 3) by finding the centroid, the frame with the highest sum of similarities, of all trajectory frames of 1 μs plain MD simulations with 2D sketches of the binding pocket and the relevant hydrogen bonds between protein and peptide on the right side, respectively. As a distance metric we used the pairwise RMSD and then used the pairwise distances to calculate a pairwise similarity.

Likely important positions of the domains

Domain	**pos 17/26**	**pos 28/35**	**pos 38/44**	**pos 71/77**	**pos 72/78**	**pos 79/85**
hPTP1E PDZ2	SER	THR	LYS	HIS	**LYS**	**ARG**
hPTP1E PDZ2 R79A	SER	THR	LYS	HIS	**LYS**	**ALA**
MAGI1 PDZ1	GLY	ASP	LYS	HIS	**ALA**	**GLN**
MAGI1 PDZ1 Q85R	GLY	ASP	LYS	HIS	**ALA**	**ARG**

TABLE 4.1: Comparison of the four studied PDZ domains regarding the residues at specific positions which where found to be important for binding. Only residues are shown, where side-chain interactions and not only backbone interactions played a role because only here the difference in the residues should matter. Special focus is on the positions 72/78 and 79/85 because these are close to SER6/GLU6/SP16/SP26 in the peptide and could therefore have a strong influence on the binding especially of peptides with charged side chains.

Q85R domains of MAGI1 were stable during the course of 1 µs plain MD simulations and the peptides remained in the binding groove. We observed unbinding events for MAGI1 PDZ1 with the EQVEAV peptide and hPTP1E PDZ2 R79A with the EQVEAV peptide and partial unbinding for MAGI1 PDZ1 with the EQVSAV peptide. We thus repeated all of these simulations once to obtain stably bound trajectories for analysis. The second simulation of hPTP1E PDZ2 R79A with the EQVEAV peptide showed a partial unbinding so we repeated it once more but obtained another partial unbinding event. We also observed a few very short partial unbinding events where we did not repeat the simulations because the short time frames of these unbindings should not influence the results too much. In the following "mean distances" is the term describing mean distances between the closest heavy atoms and they will be given as mean ± standard deviation to get a feeling how strong the residues move relative to each other and how stable the interaction is.

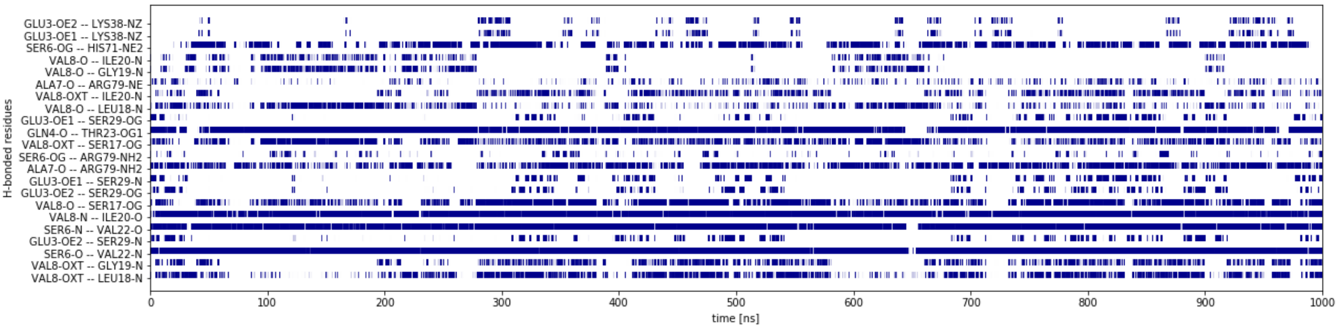
When studying the results from the individual simulations it catches the eye that hydrogen bonds between the backbone of the peptides and the backbone of the beta strand in the PDZ domain, no matter if hTPT1E PDZ2 or MAGI1 PDZ1, play the most important role in the binding, especially hydrogen bonds involving the backbone of SER6/GLU6. Additionally, the interactions between VAL8-COO$^-$-Pep and various residues in the binding pocket are usually as important. The latter are more important for the binding of the peptides to the hPTP1E PDZ2 domain and hPTP1E PDZ2 R79A domain than for the MAGI1 PDZ1 and MAG1 PDZ1 Q85R domains. It is important to mention, that the studied domains differ at positions that are likely important for the binding of charged peptides. Namely positions 79 (ARG) and 72 (LYS) in hPTP1E PDZ2 that are positions 85 (GLN) and 78 (ALA) in MAGI1 PDZ1. A focus will be on these positions in the following.

Representative snapshots of the hPTP1E PDZ2 and the MAGI1 PDZ1 domain bound to the EQVSAV peptide and EQVpSAV peptide with singly charged phosphate group are shown in fig. 4.1 and fig. 4.2. Additionally, these figures show two-dimensional sketches of the peptides inside the binding pocket and the hydrogen bonds in the system with their occupancy. The hydrogen bond occupancies and distances are listed in table 4.2 and the hydrogen bond pattern in fig. 4.3 for the hPTP1E PDZ2 domain binding the EQVSAV peptide. The tables and figures for all other simulations are shown in the supplementary material.

hPTP1E PDZ2 with EQVSAV

Interacting atoms	Occupancy [%]	Distance (STD) [Å]
SER6-O – VAL22-N	99.08	2.86 (0.21)
SER6-N – VAL22-O	96.71	3.03 (0.38)
VAL8-N – ILE20-O	96.42	2.98 (0.21)
GLN4-O – THR23-OG1	93.07	2.92 (0.97)
ALA7-O – ARG79-NH2	68.56	3.54 (1.38)
VAL8-O – SER17-OG	67.68	3.38 (1.14)
SER6-OG – HIS71-NE2	61.20	3.78 (1.31)
VAL8-OXT – SER17-OG	58.63	3.36 (0.97)
VAL8-OXT – LEU18-N	55.28	3.56 (0.92)
VAL8-OXT – GLY19-N	40.79	3.86 (1.21)
VAL8-O – LEU18-N	40.28	3.77 (1.05)
VAL8-OXT – ILE20-N	35.85	4.37 (1.28)
VAL8-O – GLY19-N	21.79	4.40 (1.24)
GLU3-OE1 – SER29-N	19.22	7.50 (3.99)
GLU3-OE2 – SER29-N	19.16	7.49 (3.99)
ALA7-O – ARG79-NE	18.24	4.10 (1.01)
GLU3-OE2 – SER29-OG	17.24	7.63 (4.48)
GLU3-OE1 – SER29-OG	17.21	7.63 (4.47)
VAL8-O – ILE20-N	17.06	5.05 (1.23)
GLU3-OE2 – LYS38-NZ	11.43	9.60 (3.76)
GLU3-OE1 – LYS38-NZ	11.36	9.57 (3.75)
SER6-OG – ARG79-NH2	10.69	5.09 (1.76)

TABLE 4.2: Percentage of hydrogen-bond occupancies and the distances between the involved heavy atoms for hydrogen bonds between hPTP1E PDZ2 and the EQVSAV peptide, existing in 10 % or more of all simulation frames. In the left column, first the peptide residue and second the protein residue is given.



(A) hPTP1E PDZ2 binding EQVSAV

FIGURE 4.3: Hydrogen bonds that exist in more than 10 % of all frames (white = no hydrogen bond, blue = hydrogen bond) between the PDZ2 domain of hPTP1E and the EQVSAV peptide.

We will first compare the complexes of the EQVSAV peptide and wildtype hPTP1E PDZ2 vs MAGI1 PDZ1. In the simulations, the EQVSAV peptide showed a similar binding behavior to both PDZ domains. In both cases the strongest interactions included backbone atoms of residues SER6-Pep and VAL8-COO$^-$-Pep (see fig. 4.1, fig. 4.2, fig. 4.3, fig. B.3, and tables 4.2 and B.7). The interactions of atoms from VAL8-COO$^-$-Pep play a weaker role when binding to MAGI1 PDZ1 than when binding to hPTP1E PDZ2. Also, the side chain of SER6-Pep has frequent interactions with hPTP1E PDZ2 but not with MAGI1 PDZ1.

Then we asked what happens for the binding of EQVSAV to hPTP1E PDZ2 when R79 is mutated to alanine. The mutation R79A in hPTP1E PDZ2 has apparently no strong influence on the binding of the EQVSAV peptide to the domain (see fig. 4.1, fig. 4.3, fig. B.2, B.6, and tables 4.2 and B.4). In both cases the hydrogen bond between the backbone oxygen atom of SER6-Pep and the backbone nitrogen atom of VAL22-Prot seems to be the most important one. Also the other backbone interactions of residues 3, 4, 6 and 8 in the peptide play important roles in both simulations. Upon the mutation R79A, the interaction of ALA7-Pep with residue 79 in the protein is lost. In both cases, the side-chain oxygen atom of SER6-Pep interacts with a side-chain nitrogen atom of HIS71-Prot even though the occupancy is reduced in the case of the mutated domain when compared to the original one. Overall in both simulations most interactions are formed between the peptide backbone and the beta strand of the PDZ domain and also between the oxygen atoms of VAL8-COO$^-$ and various residues in the binding pocket.

Next, we asked if the contacts of EQVSAV to the R79A mutant of hPTP1E PDZ2 are similar to those formed with MAGI1 PDZ1, because both do not have an arginine at position 79 (or 85 in MAGI1 PDZ1). While SER6-Pep interacts with HIS71-Prot in the hPTP1E PDZ2 domain it does not interact with HIS78-Prot in the MAGI1 PDZ1 domain, even though they are at the same position (see fig. 4.2, fig. B.2, fig. B.3, B.6, and tables B.4 and B.7). All other interactions look similar even though the specific occupancies and distances are hard to compare because of the short partial unbinding in the simulation of hPTP1E PDZ2 R79A with EQVSAV.

We now turn to the MAGI1 PDZ1 domain. What changes when EQVSAV binds either to MAGI PDZ1 or to a MAGI1 PDZ1 Q85R mutant? The overall binding modes of the EQVSAV peptide to MAGI1 PDZ1 and MAGI1 PDZ1 Q85R are quite similar (see fig. 4.2, fig. B.3, fig. B.4, fig. B.8, and tables B.7 and B.10). Notably, the mutation Q85R does not strongly alter the interaction of VAL8-COO$^-$-Pep with the protein residue at position 85. It forms hydrogens bonds with similar occupancies with GLN85-Prot in the MAGI1 PDZ1 domain as well as with ARG85-Prot in MAGI1 PDZ1 Q85R. Though, due to the mutation a new hydrogen bond between the backbone oxygen atom of ALA7-Pep and ARG85-Prot is formed.

Finally, we compare the binding of the EQVSAV peptide to this MAGI1 PDZ1 Q85R mutant where an arginine residue was introduced at the equivalent position of R79 to the binding of the peptide to the hPTP1E PDZ2 domain. In both cases, this arginine interacts with the backbone oxygen atom of ALA7-Pep. But the side chain of SER6-Pep has frequent interactions with hPTP1E PDZ2 including those with arginine and HIS71-Prot (which is HIS77-Prot in MAGI1 PDZ1) but not with MAGI1 PDZ1 Q85R (see fig. 4.1, fig. 4.3, fig. B.4, fig. B.8, and tables 4.2 and B.10).

Here, we briefly summarize the findings from the MD simulations of the EQVSAV peptide bound to various PDZ domains and variants. We found that the EQVSAV peptide binds similar to all studied PDZ domains and that mutations at position 79 in the hPTP1E PDZ2 domain and equivalently at position 85 in the MAGI1 PDZ1 domain do not strongly influence the binding. Overall, the binding of the C-terminal residue of the peptide to the respective domain seems to be more important for the binding to the hPTP1E PDZ2 domain than to the the MAGI1 PDZ1 domain.

Now we turn to the simulations of the same PDZ domains with the EQVEAV peptide. First we compare again the complexes with the two wild-type domains. While the EQVEAV peptide formed multiple interactions with ARG79-Prot when binding to the hPTP1E PDZ2 domain it did not interact at all with GLN85-Prot (similar position to ARG79) in the MAGI1 PDZ1 domain (see fig. B.1, fig. B.3, fig. B.5, fig. B.7, and tables B.1 and B.8). Overall mainly backbone interactions play a role when binding to the MAGI1 PDZ1 domain while the binding to ARG79-Prot seems to be important for the binding to the hPTP1E PDZ2 domain. ARG79-Prot play a stronger role for the EQVEAV peptide than for the EQVSAV peptide.

To see what would happen if we introduce an analogous arginine in the MAGI1 PDZ1 domain, we introduced a Q85R mutation and studied its complex with the EQVSAV peptide. The EQVEAV peptide binds in a similar manner to MAGI1 PDZ1 and MAGI1 PDZ1 Q85R but introducing an arginine at position 85 in MAGI1 PDZ1 leads to the formation of several hydrogen bonds between GLU6-Pep and ARG85-Prot (see fig. B.3, fig. B.4, fig. B.7, fig. B.8, and tables B.8 and B.11). Additionally, also ALA7-Pep interacts with ARG85-Prot.

Now we compare again the binding to the Q85R mutant of MAGI1 PDZ1 to that with the hPTP1E PDZ2 domain. The main difference in the binding modes of the EQVEAV peptide to the hPTP1E PDZ2 domain and to the MAGI1 PDZ1 Q85R domain are the hydrogen bonds between GLU6-Pep and LYS72-Prot in hPTP1E PDZ2 (see fig. B.1, fig. B.4, fig. B.5, fig. B.8, and tables B.1 and B.11). In MAGI1 PDZ1 an alanine is located at position 78 which is similar to position 72 in hPTP1E PDZ2 (see table 4.1). The backbone interactions and the hydrogen bonds with ARG79-Prot/ARG85-Prot are similar for both domains.

As mentioned before, ARG79-Prot plays a crucial role for the interaction of the hPTP1E PDZ2 domain with the peptide EQVEAV. What happens if this arginine is mutated to alanine? Upon mutating R79A in the hPTP1E PDZ2 R79A domain, multiple hydrogen bonds between ARG79-Prot and peptide residues are lost (see fig. B.1, fig. B.2, fig. B.5, fig. B.6, and tables B.1 and B.5). Furthermore, we observed a partial unbinding in the simulation of hPTP1E PDZ2 R79A with the EQVEAV peptide, which underlines the crucial role of ARG79-Prot.

Finally, we compared the binding of the EQVEAV peptide to the hPTP1E PDZ2 R79A mutant to that with the MAGI1 PDZ1 domain. As mentioned, the MAGI1 PDZ1 domain does not have a residue corresponding to ARG79-Prot. Whereas the hydrogen bonds of the backbone of GLU6-Pep of the EQVEAV peptide with the beta strand of the domain play the most important role for MAGI1 PDZ1, the interactions of VAL8-COO$^-$-Pep are more important for the binding of the peptide to the hPTP1E PDZ2 domain (see fig. B.2, fig. B.3, fig. B.6, fig. B.7, and tables B.5 and B.8). Whereas the side chain of GLU6-Pep interacts with LYS38-Prot in the hPTP1E PDZ2 domain it does not interact with any residue of the MAGI1 PDZ1 domain. Though, this could be an artifact of the partial unbinding in the simulation of hPTP1E PDZ2 R79A with the EQVEAV peptide.

Again, we briefly summarize the main findings from the MD simulations of PDZ domains and their variants with the EQVEAV peptide. We found that the binding of the EQVEAV peptide to the domains without ARG79-Prot (or ARG85-Prot) is mainly due to backbone interactions while the presence of the arginine leads to strong interactions with GLU6-Pep in the peptide. Furthermore, GLU6-Pep also binds to LYS72-Prot in hPTP1E PDZ2 that does not exist in MAGI1 PDZ1. The main difference to the EQVSAV peptide seems to be that ARG79-Prot/ARG85-Prot is much more important for the binding of the EQVEAV peptide than for the binding of the EQVSAV peptide. If the arginine is not present, mostly the peptides backbones bind to the PDZ domains.

4.3.2 Binding free energy of the EQVSAV peptide

Alchemical free energy differences by simulating bound and free state separately

Next, we performed alchemical simulations to obtain the binding free energy difference between the EQVSAV and EQVEAV peptides in complex with the hPTP1E PDZ2 domain which has also been studied by Toto et al. [156]. The experimental binding free energy difference is

$$\Delta\Delta G_{\text{bind}}^{\text{exp,EQVEAV}-\text{EQVSAV}} = 6.19 \pm 0.73\,\text{kJ}\,\text{mol}^{-1} \tag{4.2}$$

at a NaCl concentration of $n_{\text{NaCl}} = 150\,\text{mmol}$. This alchemical transformation of a peptide involves a charge change since glutamic acid is most commonly found in the form of glutamate at physiological pH which has a total charge of -1. Glutamic acid is thus a good system to study the effect of charge changes since the common force fields are usually more reliable for natural amino acid side chains compared to the ones with post-translational modifications such as phosphorylation that are also relevant for peptide binding to PDZ domains. The resulting binding free energy difference from an alchemical simulation where bound and free states simulated separately is shown in table 4.5. The difference between the simulation result and the one from experiment is $1.55\,\text{kJ}\,\text{mol}^{-1}$ which is a very small deviation for common binding free energy methods. This suggests that the methodology works as expected also for alchemical simulations where the total system charge is altered during the course of the simulation.

Absolute binding free energy using PaCS-MD and Markov state models

We also performed PaCS-MD simulations to obtain the absolute binding free energy for the EQVSAV peptide binding to the hPTP1E PDZ2 domain. Fig. 4.4 shows the inter-COM distance between the hPTP1E PDZ2 domain and the EQVSAV peptide as a function of the number of cycles in the PaCS-MD trials by plotting always the largest inter-COM distance in each cycle. From these plots we estimated the distance ranges of the bound (inter-COM distance < 1.4 nm), partially-bound (1.4 nm < inter-COM distance < 2.2 nm) and unbound (2.2 nm < inter-COM distance) states. The PaCS-MD simulation needed about 27 ± 8 cycles corresponding to 40 ± 12 ns of simulation time on average to completely dissociate the protein and the peptide. For the analysis of the simulations we built MSMs using the 3D-COM positions of the peptides as a feature set similar to the work on dissociation PaCS-MD by Tran and Kitao [27].

The absolute binding free energy resulting from PaCS-MD for the unphosphorylated EQVSAV peptide binding to the hPTP1E PDZ2 domain is

$$\Delta G_{\text{bind,PaCS}}^{\text{EQVSAV}} \approx -26.68\,\text{kJ}\,\text{mol}^{-1} \tag{4.3}$$

for a MSM lag time of 40 ps which nearly perfectly matches the experimental binding free energy $\Delta G_{\text{bind,exp.}}^{\text{EQVSAV}} = -27.86\,\text{kJ}\,\text{mol}^{-1}$ [156].

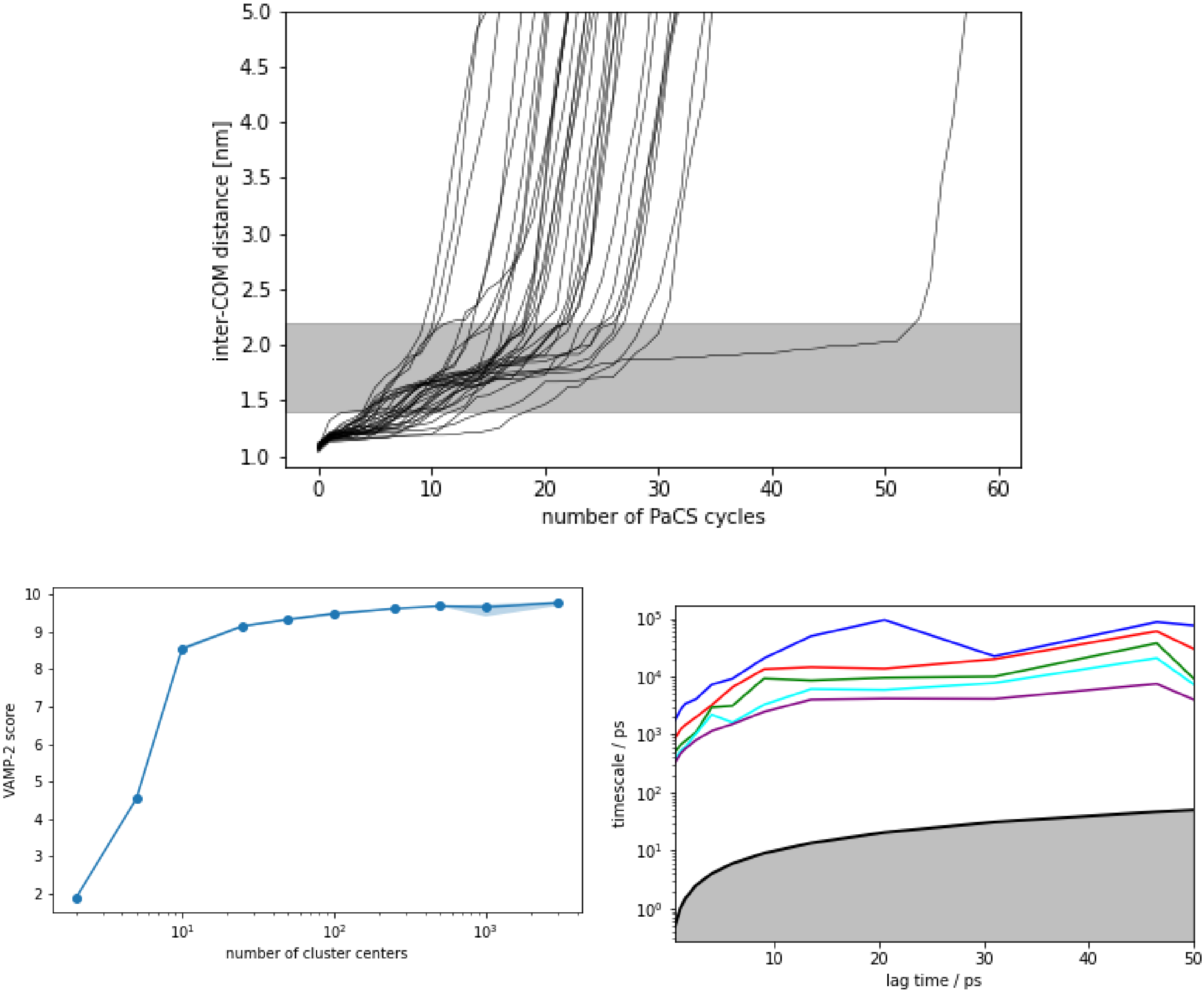


FIGURE 4.4: Analysis of the PaCS-MD simulation of the EQVSAV peptide undergoing stimulated dissociation from the hPTP1E PDZ2 domain. Top: Inter-COM distance as a function of the number of PaCS-MD cycles. The inter-COM distance range is divided in the bound, partially-bound and unbound regions. Bottom left: VAMP-2 scores for as a function of the number of cluster centers. The analysis was repeated five times for every studied number of cluster centers and the standard deviation is shown as light-blue area. Bottom right: Implied MSM time scales as a function of the lag time τ used.

4.3.3 Unbiased molecular dynamics of PDZ domains binding the phosphorylated EQVpSAV peptide

Most of the performed plain MD simulations of the EQVpSAV peptides bound to the PDZ2 and PDZ2 R79A domains of hPTP1E and to the PDZ1 and PDZ1 Q85R domains of MAGI1 were stable during the course of 1 µs plain MD simulations and the peptides remained in the binding groove. We observed partial unbinding events for MAGI1 PDZ1 Q85R with the EQVpSAV peptide and thus repeated this simulation once but obtained another trajectory with partial unbinding.

For the hPTP1E PDZ2 domain, we compared the binding of the two versions of the EQVpSAV peptide, namely with a singly charged phosphate group and with a doubly charged phosphate group. The hPTP1E PDZ2 domain shows similar very frequent hydrogen bonds when binding the EQVpSAV peptide either with singly or doubly charged phosphate (see fig. 4.1, fig. B.1, fig. B.5 and tables B.2 and B.3). These are the backbone interactions between VAL8-Pep and ILE20-Prot, SP16-Pep and VAL22-Prot and the interaction of the backbone of ALA7-Pep with the side chain of ARG79-Prot. In both simulations, the phosphate group interacts with the nitrogen atoms in the side chain of ARG79-Prot but the mean distances between the phosphate oxygen atoms and the nitrogen atoms of the ARG79-Prot side chain are shorter on average in the simulation with a doubly charged phosphate compared to the one with a singly charged phosphate. The effect is even stronger for the distances between the phosphate oxygen atoms and the side-chain nitrogen atoms of LYS72-Prot.

As will be described below, the binding thermodynamics of the peptide with doubly charged phosphate deviated strongly from the experimental value. Hence, only the singly charged phosphate variant was used in the remaining MD simulations.

First, we compared binding of unphosphorylated and phosphorylated variants of EQVSAV to the hPTP1E PDZ2 domain (see fig. 4.1, fig. 4.3, fig. B.1, and tables 4.2 and B.2). When the phosphate group is added to the EQVSAV peptide, the interaction between the backbone nitrogen atom of SER6-Pep/SP16-Pep and the backbone oxygen atom of VAL22-Prot in hPTP1E PDZ2 is lost and the occupancy of the interaction between the backbone oxygen atom of SER6-Pep/SP16-Pep with the backbone nitrogen atom of VAL22-Prot is reduced. Additionally, the latter interaction shows a larger standard deviation for the distance between the atoms for the binding of the EQVpSAV peptide suggesting, that the interaction is more flexible in this case. Whereas the side-chain oxygen atom of SER6-Pep in EQVSAV is interacting with HIS71-Prot, the phosphate oxygen atoms of EQVpSAV are interacting with ARG79-Prot and LY72-Prot instead. ARG79-Prot in the hPTP1E PDZ2 domain usually forms a hydrogen bond with the backbone oxygen atom of ALA7-Pep. This interaction is populated even more often when the peptide is phosphorylated. Therefore it seems that the phosphorylation of SER6-Pep stabilizes the interaction of the peptide with ARG79-Prot. The hydrogen bond between the backbone nitrogen atom of VAL8-COO^- in the peptide and the backbone oxygen atom of ILE20 in the protein is one of the most important interactions for the binding of the peptides to the protein, independent of the phosphorylation state, even though other interactions, for example the one between the backbone oxygen atom of SER6/SP16/SP26 in the peptide and the backbone nitrogen atom of VAL22 in the protein, are as important for the binding.

Next, we analyzed whether the glutamate-containing peptide EQVEAV peptide is a good mimic of the EQVpSAV peptide when bound to hPTP1E PDZ2. Even though EQVEAV and EQVpSAV with singly charged phosphate both carry a single negative charge, they form different interactions with the hPTP1E PDZ2 domain. In particular, the side chains of GLU6-Pep in EQVEAV and SP16-Pep in EQVpSAV show different interactions (see fig. 4.1, fig. B.1, fig. B.5 and tables B.1 and B.2). Whereas the phosphate oxygen atoms of SP16-Pep are binding repeatedly to the nitrogen atoms of the ARG79-Prot side chain with specific interactions having an occupancy of nearly 50 %, leading to a continuous interaction between SP16-Pep and ARG79-Prot, the side-chain oxygen atoms of GLU6-Pep bind to ARG79-Prot only sometimes and more often to LYS72-Prot. SP16-Pep binds to LYS72-Prot with a similar occupancy as GLU6-Pep.

As mentioned before, ARG79-Prot plays a crucial role for the interaction of the hPTP1E PDZ2 domain with the peptide EQVEAV but less for the interaction with the EQVSAV peptide. Thus we wanted to compare the interactions of the hPTP1E PDZ2 R79A mutant with the EQVSAV and EQVEAV peptides to the interactions with the EQVpSAV peptides (see fig. B.2, B.6, B.7, and tables B.4, B.5, and B.6). Whereas the side chain of SER6-Pep of the EQVSAV peptide forms a hydrogen bond with HIS71-Prot, SP16-Pep of the EQVpSAV peptide forms hydrogens bonds with LYS38-Prot and LYS72-Prot. Upon the mutation R79A no interactions are formed with residue 79 of the domain. Due to the partial unbinding in the simulation of hPTP1E PDZ2 R79A binding the EQVEAV peptide, the results are hard to compare. Though, it is interesting to see that in both simulations, with the EQVEAV peptide and with the EQVpSAV peptide, GLU6-Pep/SP16-Pep form hydrogen bonds with LYS38-Prot.

Now we turn to the simulations with the MAGI1 PDZ1 domain, that also does not have an arginine at position 79/85 like the hPTP1E PDZ2 R79A mutant. Comparing the simulations of the EQVpSAV peptide binding these two domains we see that the hydrogen bonds between the backbone oxygen atom of SP16-Pep and a backbone nitrogen atom in the beta sheet of the protein, as well as between various atoms of VAL8-COO$^-$-Pep and the protein seem to be similarly important for the binding of the EQVpSAV peptide with singly charged phosphate (see fig. 4.2, fig. B.2, B.3, B.7, and tables B.6 and B.9). The binding modes differ mostly in the interactions between the phosphate oxygen atoms of SP16-Pep and backbone nitrogen atoms of LYS38-Prot and LYS72-Prot in hPTP1E PDZ2 R79A which have no similar interactions when the peptide binds to MAGI1 PDZ1 even though MAGI1 PDZ1 also has a lysine (LYS44-Prot) at the same position as LYS38-Prot in hPTP1E PDZ2.

While it is interesting to study the different binding behaviour of the EQVpSAV peptide to different domains, it is also interesting to study the differences between the binding of the EQVSAV, EQVEAV and EQVpSAV peptides to the MAGI1 PDZ1 domain (see fig. 4.2, B.3, B.7, and tables B.7, B.8, and B.9). The hydrogen bond between the backbone oxygen of SER6-Pep/SP16-Pep and the backbone nitrogen of VAL31-Prot seems to be similarly important for MAGI1 PDZ1 binding to the EQVSAV as for the EQVpSAV peptide with singly charged phosphate. On the other hand, the interaction between the backbone nitrogen atom of SER6-Pep/SP16-Pep and the backbone oxygen atom of VAL31-Prot is lost for SP16-Pep compared to SER6-Pep, similarly to the hPTP1E PDZ2 domain binding the EQVSAV and EQVpSAV peptide with singly charged phosphate. Overall the interactions of the MAGI1 PDZ1 domain with the EQVSAV peptide and with the EQVpSAV peptide are hard to compare because the simulation of MAGI1 PDZ1 binding EQVpSAV peptide showed a partial unbinding. Similarly, specificities of the binding of the phosphorylated EQVpSAV peptide with singly charged phosphate and the EQVEAV peptide to the MAGI1 PDZ1 domain are hard to compare due to a partial unbinding in the simulation with the EQVpSAV peptide though the binding looks similar. One important result to mention is that in both simulations the side chains of GLU6-Pep/SP16-Pep do not show any hydrogen bonds with any protein residue.

As before, we also introduced a Q85R mutation in MAGI1 PDZ1 and studied its complex with the EQVpSAV peptide. The mutation Q85R in MAGI1 PDZ1 introduces interactions between the phosphate oxygen atoms of SP16-Pep of the EQVpSAV peptide with singly charged phosphate and the ARG85-Prot side-chain nitrogen atoms even though they are not very frequent. The hydrogen bonds between VAL8-COO^--Pep and the MAGI1 PDZ1 domain seem to be weakened in the simulation with the mutation compared to the simulation without mutation (see fig. 4.2, B.3, B.4, B.9, and tables B.9 and B.12). This simulation of the EQVpSAV peptide binding the MAGI1 PDZ1 Q85R domain showed a partial unbinding in two simulations wherefore it is hard to compare the results with the results for the binding of the EQVSAV and EQVEAV peptides to the same domain.

Similarly, it is hard to compare the binding of the EQVpSAV peptide to the hPTP1E PDZ2 domain and to the MAGI1 PDZ1 Q85R mutant that both have an arginine at position 79 (or equivalently 85 in MAGI1 PDZ1). Though, there it is apparent, that the peptide binds much more tightly to the hPTP1E PDZ2 domain and moves much stronger in the binding pocket of MAGI1 PDZ1 Q85R which could be an artifact of the partial unbinding of the N-terminal tail when binding to MAGI1 PDZ1 Q85R. All interactions of the phosphate oxygen atoms of SP16-Pep have higher occupancies when binding to the hPTP1E PDZ2 domain (see fig. 4.1, B.1, B.4, B.9, and tables B.2 and B.12).

In summary, the MD simulations of PDZ domains and variants complexed with phosphorylated peptides led to the following general findings. The overall binding behaviour is similar. The binding is mainly mediated by backbone-backbone interactions. When an arginine is present at position 79 (or 85 in the MAGI1 PDZ1 Q85R domain) or when a lysine is present at position 72 in the hPTP1E PDZ2 domain and its variants, the glutamate and the phosphate both form hydrogen bonds with these residues. The presence of the phosphate also increases the occupancy of a hydrogen bond between the backbone oxygen of ALA7-Pep with ARG79-Prot. Overall it seems that the phosphorylation of the EQVSAV peptide leads to a weaker binding to MAGI1 PDZ1 because we had multiple simulations of the EQVpSAV peptide with MAGI1 PDZ1 and MAGI1 PDZ1 Q85R showed unbinding or partial unbinding. Furthermore, the interactions between the glutamate and the protein and between the phosphate and the protein appear to be different, at least in their occupancies. Therefore, glutamic acid is likely not a good mimic for phosphorylation in the EQVSAV peptide.

4.3.4 Binding free energy of the phosphorylated EQVpSAV peptide binding to PDZ domains

Alchemical free energy differences using the "Double system in a single box" setup

At first we simulated the alchemical change from the EQVSAV peptide to the EQVpSAV peptide binding to the hPTP1E PDZ2 domain using the "double system in a single box setup". The phosphate group is usually found in the dibasic form (–OPO_3^{-2}, called SP2 in the CHARMM force field) at physiological pH (phosphoserine $pK_a = 5.6$ [356]). Since our simulations were run at $pH = 7.2$, we used this protonation state for our first simulations. In the following the index SP2 will be used to label results for the peptide EQVpSAV with the phosphate group in the dibasic form

	$\Delta\Delta G_{\mathrm{bind}}$ [kJ mol^{-1}]		
n_{NaCl} [mmol]	exp.	alch. sim. (SP1)	alch. sim. (SP2)
0	3.79 ± 0.28	-4.25 ± 0.56	
150	4.65 ± 0.35	-4.68 ± 0.48	-15.94 ± 1.68
300	5.57 ± 0.35	-4.12 ± 0.65	

TABLE 4.3: Binding free energy differences between EQVpSAV/EQVSAV peptides binding to hPTP1E PDZ2 domain obtained by "double system in a single box" alchemical simulations in comparison with experiment at 283 K.

while SER will be used for the unphosphorylated EQVSAV peptide. The resulting binding free energy difference between the phosphorylated and unphosphorylated states using a salt concentration in the simulation of $n_{\mathrm{NaCl}} = 150\,\mathrm{mmol}$ was

$$\Delta\Delta G_{\mathrm{bind}}^{\mathrm{sim,SP2-SER}} = -15.94 \pm 0.17\,\mathrm{kJ\,mol^{-1}} \tag{4.4}$$

which favors binding of the phosphorylated peptide. This did not reproduce the experimental value of $\Delta\Delta G_{\mathrm{bind}}^{\mathrm{exp,phosphorylated-unphosphorylated}} = 4.65 \pm 0.35\,\mathrm{kJ\,mol^{-1}}$ [156] at the same salt concentration which favors binding of the unphosphorylated peptide. The experimental results were calculated from the experimental K_{D} values reported by Toto et al. [156] using $\Delta G = RT \ln \frac{K_{\mathrm{D}}}{C^0}$ with the standard concentration $C^0 = 1M$. In this manuscript all presented relative binding free energy differences are always the absolute binding free energy of the phosphorylated peptide minus the absolute binding free energy of unphosphorylated peptide $\Delta\Delta G_{\mathrm{bind}} = \Delta G_{\mathrm{bind}}^{\mathrm{phosphorylated}} - \Delta G_{\mathrm{bind}}^{\mathrm{unphosphorylated}}$.

In order to test the influence of the phosphate protonation state onto the results we additionally simulated the monobasic form (–OPO_3H^-, called SP1 in CHARMM force field) using the same setup and salt concentration. In the following the index SP1 will be used to label results for the peptide EQVpSAV with the monobasic phosphate group. The result was

$$\Delta\Delta G_{\mathrm{bind}}^{\mathrm{sim,SP1-SER}} = -4.68 \pm 0.30\,\mathrm{kJ\,mol^{-1}} \tag{4.5}$$

which is of similar absolute magnitude but of opposite sign than the experimental result. Such a deviation between experimental and simulated results of $9.33\,\mathrm{kJ\,mol^{-1}} \approx 2.23\,\mathrm{kcal\,mol^{-1}}$ is a common size of error ranges in binding free energy simulations involving charge changes and flexible ligands. The different sign could thus be an artifact due to a similar size of common free energy error and the binding free energy difference between phosporylated and unphosphorylated peptide.

In order to understand the influence of the ionic concentration onto the alchemical binding free energy simulations we further simulated the systems at other concentrations of NaCl. The results are shown in table 4.3. In the experiments, a linear increase of the salt concentration linearly shifts the free energy difference in the direction favouring the binding of the unphosphorylated peptide. At higher salt concentration, the phosphorylated form can be more favorably coordinated in the solvent.

As shown in table 4.4, mutating ARG79-Prot to alanine destabilizes binding of the phosphorylated peptide as can be expected. The MAGI1 PDZ1 domain actually showed an even more pronounced preference for the unphosphorylated peptide.

PDZ domain	$\Delta\Delta G_{\mathrm{bind}}$ [kJ mol^{-1}] alch. sim. (SP1)	alch. sim. (SP2)
hPTP1E PDZ2	-4.68 ± 0.48	-15.940 ± 1.68
hPTP1E PDZ2 R79A	1.33 ± 0.46	
MAGI1 PDZZ1	7.66 ± 0.57	

TABLE 4.4: Comparison of binding free energy differences between EQVpSAV/EQVSAV peptides binding to hPTP1E PDZ2, hPTP1E PDZ2 R79A and MAGI1 PDZ1 obtained by "double system in a single box" alchemical simulations at 283 K and 150 mmol NaCl.

	$\Delta\Delta G_{\mathrm{bind}}$ [kJ mol^{-1}] uncorrected	standard state corr.	elec. corr.	corrected
alch. sim. (GLU)	6.38	-0.03	-1.71	4.64
alch. sim. (SP1)	-4.34 ± 0.90	-1.91576 ± 0.00009	-2.11 ± 0.64	-8.36 ± 1.54
alch. sim. (SP2)	-7.94	-1.92	-4.69	-14.55

TABLE 4.5: Binding free energy differences between EQVpSAV/EQVSAV peptides as well as between EQVEAV/EQVSAV peptides binding to hPTP1E PDZ2 domain obtained by simulating alchemical changes of bound and free state separately at 283 K and $n_{\mathrm{NaCl}} = 150$ mmol with standard state and electrostatic corrections similar to our previous study [32]. The alchemical simulations for the change between SER and SP1 were repeated five times and a statistical error could therefore be estimated. The perturbations between SER and SP2 and the one between SER and GLU were only performed once.

Alchemical free energy differences by simulating bound and free state separately

The results from the alchemical simulations using the "double system in a single box" setup had the wrong sign. This was supposedly due to a too high error of the alchemical simulation binding free energies due to the force field which did not correctly reproduce the differential stabilization of phosphoserine in the protein vs. in solvent and possibly also due to the methodology and the non-existent possibility to use error corrections for results from "double system in a single box" setups. In order to test how the methodology influences the results and if the established correction terms of another methodology improve the results we also performed alchemical simulations of the bound and free states separately and applied the standard state and electrostatic finite-size corrections similar to our previous study [32]. The resulting binding free energy differences and corrections are listed in table 4.5. Using this methodology the binding free energy difference between the phosphorylated EQVpSAV peptide with a monobasic phosphate group and the unphosphorylated EQVSAV peptide deviated even further from the experimental value than the one from the "double system in a single box setup". The difference between the monobasic and dibasic forms of the phosphate was smaller compared to the latter method but was still large and simulations with the dibasic form also did not match the experimental binding free energy difference.

To validate that the wrong sign we obtained for the free energy difference between the phosphorylated and unphosphorylated versions of the EQVSAV peptide

	$\Delta\Delta G_{\mathrm{bind}}$ [kJ mol^{-1}]			
	uncorrected	standard state corr.	elec. corr.	corrected
HSD	-4.34 ± 0.90	-1.91576 ± 0.00009	-2.11 ± 0.64	-8.36 ± 1.54
HSE	-7.11	-1.92	-2.29	-11.32
HSP	-10.11 ± 3.36	-1.92	-3.13	-14.49 ± 3.36

TABLE 4.6: Binding free energy differences between EQVpSAV/EQVSAV peptides binding to hPTP1E PDZ2 domain with different protonation states of HIS71 obtained by simulating alchemical changes of bound and free states separately at 283 K and $n_{\mathrm{NaCl}} = 150\,\mathrm{mmol}$ with standard state and electrostatic corrections similar to our previous study [32]. The monobasic form (SP1) of phosphoserine was used. The alchemical simulations were repeated five times and a statistical error could therefore be estimated except for the one with HSE which was performed only once.

was not an artifact due to an incorrect protonation state of HIS71 in the binding pocket of hPTP1E, we performed alchemical simulations for the change between EQVSAV and EQVpSAV with the monobasic form of phosphoserine for all three possible protonation states of this residue. The proton can be either on the atom ND1 (HSD), on NE2 (HSE) or on both (HSP). Usually the protonation was set to HSD using PROPKA3 [364, 365] as mentioned in the structure preparation section. The results are shown in table 4.6. Modifying the histidine protonation increases the difference between the experimental and simulation results and is therefore also not the reason for the overall difference.

Another reason for the difference between experiment and simulation could be the speed at which the alchemical simulations were performed. Therefore we performed one simulation for the change between EQVSAV and EQVpSAV with monobasic phosphoserine and histidine protonation HSD at $T = 283\,\mathrm{K}$ and $n_{\mathrm{NaCl}} = 150\,\mathrm{mmol}$ with a switching time of 10 ns instead of 1 ns. The resulting free energy difference was

$$\Delta\Delta G_{\mathrm{bind}}^{\mathrm{sim10ns,SP1-SER}} = -8.62\,\mathrm{kJ\,mol^{-1}} \tag{4.6}$$

including all correction terms which is very similar to the simulations with a switching time of 1 ns ($-8.36\,\mathrm{kJ\,mol^{-1}}$). Thus we conclude that the switching time was chosen long enough to obtain converged binding free energy differences from the alchemical simulations.

Absolute binding free energy of phosphorylated EQVpSAV peptide using PaCS-MD and Markov state models

Fig. 4.4 shows the inter-COM distance between the hPTP1E PDZ2 domain and the EQVpSAV peptide as a function of the number of cycles in the PaCS-MD trials by plotting always the largest inter-COM distance in each cycle. From these plots we estimated the bound (inter-COM distance < 1.5 nm), partially-bound (1.5 nm < inter-COM distance < 2.2 nm) and unbound (2.2 nm < inter-COM distance) states. For the unphosphorylated peptide the bound state ranged up to 1.4 nm. The PaCS-MD simulation for hPTP1E PDZ2 and the phosphorylated EQVpSAV with singly charged (SP1) phosphate needed 45 ± 46 cycles corresponding to 67 ± 68 ns simulation time on average to completely dissociate the peptide from the protein. For the analysis of the simulations we built MSMs using the 3D-COM positions of the peptides as a feature set similar as has was for dissociation PaCS-MD by Tran and Kitao [27].

The absolute binding free energy resulting from PaCS-MD for the phosphorylated EQVpSAV peptide with singly charged (SP1) phosphate binding the hPTP1E PDZ2 domain at a lag time of 40 ps is

$$\Delta G_{\text{bind,PaCS}}^{\text{EQVpSAV(SP1)}} \approx -40.68\,\text{kJ}\,\text{mol}^{-1} \tag{4.7}$$

which is also in the same order of magnitude as the experimental binding free energy $\Delta G_{\text{bind,exp.}}^{\text{EQVpSAV(SP1)}} = -23.21\,\text{kJ}\,\text{mol}^{-1}$ [156] but has a stronger deviation from the experimental value

$$\Delta\Delta G_{\text{bind}}^{\text{experiment}-\text{PaCS,SP1}} = 17.47\,\text{kJ}\,\text{mol}^{-1} \tag{4.8}$$

than the result for the unphosphorylated peptide. The resulting calculated binding free energy difference between the binding of the phosphorylated EQVpSAV peptide and the unphosphorylated EQVSAV peptide is

$$\Delta\Delta G_{\text{bind}}^{\text{PaCS,SP1}-\text{SER}} = -14.00\,\text{kJ}\,\text{mol}^{-1} \tag{4.9}$$

which again incorrectly favors binding of the phosphorylated peptide as was found by the non-equilibrium alchemical perturbation simulations.

Therefore it is likely that the deviations of the alchemical simulations of systems with phosphorylated serines are due to unoptimized force field parameters for phosphoserine.

4.4 Discussion

In this work we wanted to study how peptides bind to PDZ domains and how this is influenced by mutation in the PDZ domains as well as by mutation and phosphorylation in the peptide. We studied several systems, namely the hPTP1E PDZ2 domain and the mutated domain hPTP1E PDZ2 R79A as well as the MAGI1 PDZ1 domain and the mutated domain MAGI1 PDZ1 Q85R all binding the EQVSAV, EQVEAV and phosphorylated EQVpSAV peptides with singly charged phosphate. Additionally, we studied the binding of the phosphorylated EQVpSAV peptide with doubly charged phosphate to the hPTP1E PDZ2 domain to get insight into the influence of the phosphate charge state on the binding.

Based on the simulations we identified the following general features of motif 1 peptides binding to PDZ domains. Overall, they bind mostly due to backbone-backbone interactions with the beta strand in the binding pocket of the PDZ domains. Most prominently the backbone of the residue at position -2 from the C terminus (serine, phosphoserine or glutamic acid) forms very stable backbone interactions. Additionally, the hydrophobic C-terminal residue of the peptide (here valine), usually present as a charged C-terminus at physiological pH, forms strong hydrogen bonds with surrounding residues of the PDZ domain (residues 17-20 in hPTP1E PDZ2 and residues 27-29 in MAGI1 PDZ1). It is interesting to note, that the former effect is stronger for the MAGI1 PDZ1 domain while the latter is more important when binding to the hPTP1E PDZ2 domain but both are relevant in both cases. Overall, most of the simulations with MAGI1 PDZ1 or MAGI1 PDZ1 Q85R showed dissociation or partial unbinding. Therefore we conclude that the binding of the peptides is overall much weaker to the MAGI1 PDZ1 domain than to the hPTP1E

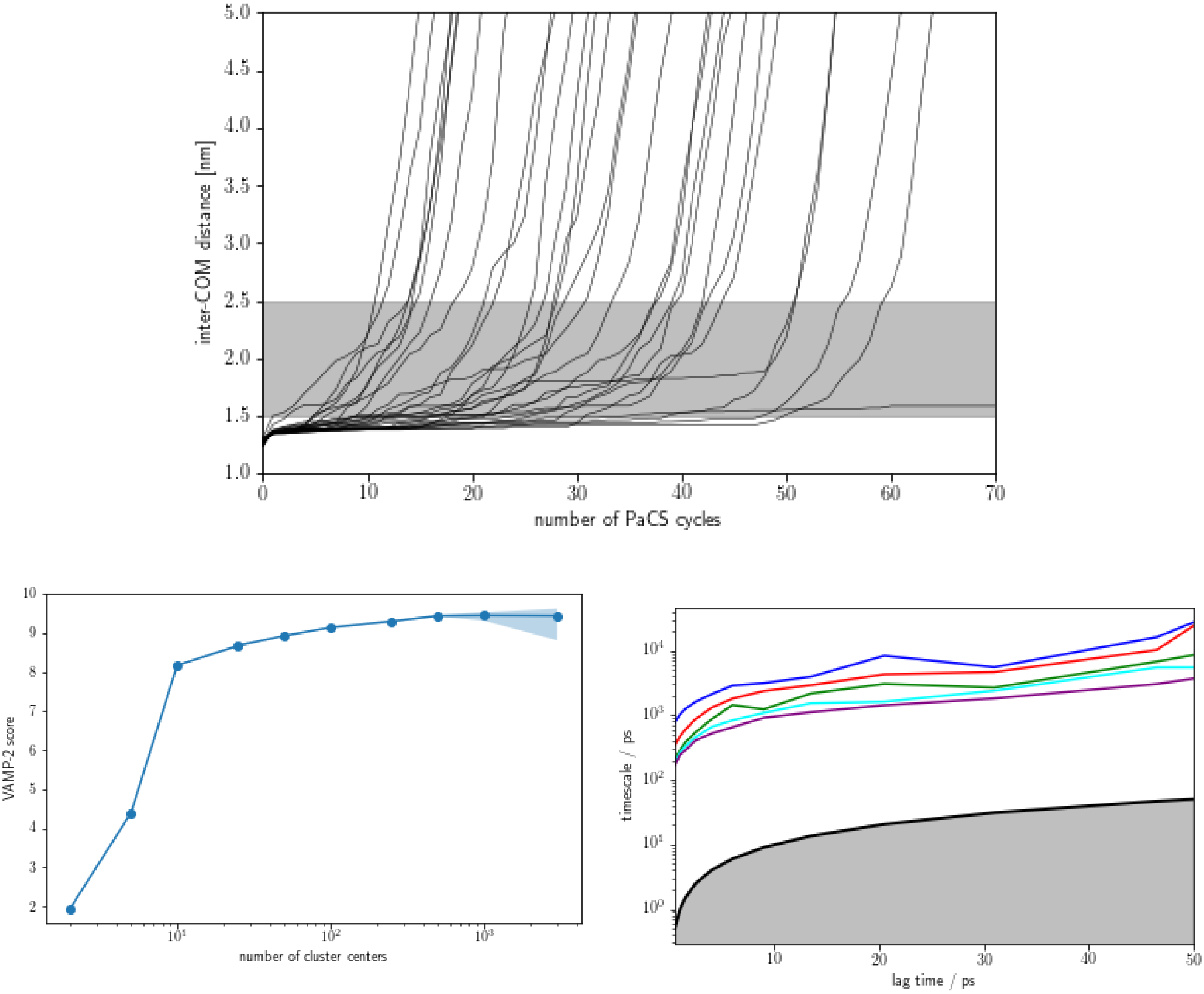


FIGURE 4.5: Analysis of the PaCS-MD simulation of the EQVpSAV peptide undergoing stimulated dissociation from the hPTP1E PDZ2 domain. Top: Inter-COM distance as a function of the number of PaCS-MD cycles. The inter-COM distance range is divided in the bound, partially-bound and unbound regions. Bottom left: VAMP-2 scores for as a function of the number of cluster centers. The analysis was repeated five times for every studied number of cluster centers and the standard deviation is shown as light-blue area. Bottom right: Implied MSM time scales as a function of the lag time τ used.

PDZ2 domain where much fewer of these events were observed. Maybe this results from the weaker binding of the C-terminal residue in MAGI1 PDZ1.

It is important to mention that our simulations were always started from bound states and we had to create several structures by exchanging the serine for glutamate or by adding the phosphate group. Even though equilibration simulations were performed in each case, it is not certain whether the simulation time was sufficient for the PDZ domains to adapt to the change in the peptide. This could influence the results because results from Morra et al. for the PDZ2 domain in complex with the RAGEF2 C-terminal peptide with sequence ENEQVSAV (similar to our system) showed that the core of the PDZ2 domain rearranged upon ligand binding in the region between beta strand 2 and loop L23 [359].

Furthermore, we were interested in how the change of serine to glutamate or phosphoserine affects the binding of the peptide to the PDZ domains. From unbiased simulations with these charged peptides, we realized that glutamate as well as phosphoserine like to bind to ARG79 in the hPTP1E PDZ2 domain. Therefore we introduced the mutation R79A in the hPTP1E PDZ2 domain and performed additional plain MD simulations with this mutated domain. Because we also observed hydrogen bonds of these charged residues to LYS72, we additionally wanted to study the influence of this residue on the binding of the peptides. We searched for another PDZ domain which we could use as a study system and found MAGI1 PDZ1 that does not contain an arginine at position 85 that maps onto position 79 in hPTP1E PDZ2. Instead, MAGI1 PDZ1 has a glutamine and additionally has an alanine instead of a lysine at the position 78 that maps onto position 72 in hPTP1E PDZ2. To have the possibility to study all four cases, i. e. a PDZ domain having lysine at position 72 and arginine at position 79 (positions in the hPTP1E PDZ2 domain), a domain having only the former, a domain having only the latter and a domain that has none of both, we also introduced the mutation Q85R in the MAGI1 PDZ1 domain. All of these domains also contain a lysine at position 38 in hPTP1E PDZ2 or 44 in MAGI1 PDZ1 (which is the same structural position).

From the plain MD simulations, we see that the glutamate of the EQVEAV peptide as well as the phosphate group of the EQVpSAV peptide interact repeatedly with ARG79-Prot (or ARG85-Prot in MAGI1 PDZ1 Q85R) if it is available. Additionally, both also repeatedly form hydrogen bonds with LYS72-Prot (only in hPTP1E PDZ2 and hPTP1E PDZ2 R79A) when it is available.

To mechanistically characterize the strength of peptide binding to PDZ domains we performed multiple alchemical simulations to obtain relative binding free energy differences as well as two parallel cascade selection molecular dynamics (PaCS-MD) simulation to obtain the absolute binding free energies for the EQVSAV peptide and the EQVpSAV peptide with singly charged phosphate to the hPTP1E PDZ2 domain. We obtained a reasonable result close to the experimental result for the binding free energy difference between the EQVEAV peptide and the EQVSAV peptide, suggesting that the procedure worked as expected. Inconveniently, the calculated binding free energy difference between the phosphorylated EQVpSAV peptide and the EQVSAV peptide has the wrong sign, i. e. even though the experimental results suggests that the phosphorylated peptide should bind weaker to the hPTP1E PDZ2 domain it bound stronger in our simulations. We performed simulations for EQVpSAV peptide with either singly or doubly charged phosphate to see if the phosphorylation state is the reason for this deviation. However, the deviation was even larger for the alchemical simulations with the EQVpSAV peptide with doubly charged phosphate. To understand if the deviation is an artifact of the setup of the alchemical simulations we performed them in two ways, using the so-called “double system in a singly box"-setup and by performing both the free and bound state separately and then using standard state and charge corrections. The deviations were similar for both setups. Furthermore, we wanted to rule out that this deviation is a general artifact of alchemical simulations, even though this is not likely because the alchemical simulations for the difference between the EQVEAV peptide (that is also charged) and the EQVSAV peptide worked well. Hence, we also performed PaCS-MD simulations to determine the absolute binding free energies for the EQVSAV. Here, we obtained a reasonable result close the experimental result for the EQVSAV peptide. Using the same setup for the PaCS-MD simulations with the EQVpSAV peptide we obtained a similar difference between our and the experimental result as we obtained from the alchemical simulations. This suggests that the issue is not with the setup of our simulations but likely due to nonoptimal force field parameters of the phosphate group or especially phosphoserine in the CHARMM36m force field. This is backed by results for the binding of phosphorylated peptides to 14-3-3 proteins which we recently studied [32]. In this study, we analyzed the binding of phosphorylated and unphosphorylated peptides to the 14-3-3η protein. 14-3-3 proteins usually bind only phosphorylated peptides and the binding of their unphosphorylated counterparts is too weak to be measured experimentally [104, 338, 339]. Also for that system we performed alchemical simulations where we simulated the bound and the free state separately. Considering, that one would expect absolute binding free energies for the binding of the unphosphorylated forms of nearly $0\,\mathrm{kJ\,mol^{-1}}$, the resulting binding free energy difference obtained from the alchemical simulations should be close to the absolute binding free energy of the phosphorylated form. Following this path, a deviation of around $6-8\,\mathrm{kJ\,mol^{-1}}$ from the experimental value was observed. This deviation has a similar size as the deviation of the relative binding free energy difference between the phosphorylated EQVpSAV and unphosphorylated EQVSAv peptide binding the hPTP1E PDZ2 domain from the experimental difference and also the deviation of the absolute binding free energy difference of the phosphorylated EQVpSAV peptide to the hPTP1E PDZ2 domain from the experimental value obtained by PaCS-MD. Additionally, another recent study also suggested that recent force fields still "inconsistently portray the microscopic details of phosphorylation" [370]. There, also a deviation of around $8\,\mathrm{kJ\,mol^{-1}}$ from the experimental result was found. Panel et al. also experienced errors in the range between $1-3\,\mathrm{kcal\,mol^{-1}} \approx 4-13\,\mathrm{kJ\,mol^{-1}}$ in their alchemical simulation for C-terminal peptides binding to the TIAM1 PDZ domain when using an additive force field [360]. They suggested that at least around $0.8\,\mathrm{kcal\,mol^{-1}}$ were the result of a systematic error.

Because the experimental binding free energy difference between the EQVpSAV peptide and the EQVSAV peptide binding to the hPTP1E PDZ2 domain is only around $4\,\mathrm{kJ\,mol^{-1}}$, an error of this size, i. e. for an error of around double the actual experimental value, could explain the wrong sign. If we assume that the phosphate parameters overestimate the interactions, the unbiased MD simulations are also influenced by these deviations. It is possible that salt bridges between the phosphate group and other charged residues are overestimated. Another possibility is that the coordination of the phosphate group in water in our simulations is less favorable than in a real system. Overall, we expect these deviations to influence the binding of the peptide but we feel that the general effects are still comparable. The errors in the alchemical simulations should be similar for all the ones involving a phosphate group and thus we can still compare them to each other.

The mutation R79A in the hPTP1E PDZ2 domain shifted the binding free energy difference between the EQVpSAV peptide with singly charged phosphate and the

EQVSAV peptide by around 6 kJ mol^{-1}, i. e. the phosphorylated peptide binds much stronger to the hPTP1E PDZ2 domain than to the hPTP1E PDZ2 R79A domain. Similarly, the phosphorylated peptide binds much stronger to the hPTP1E PDZ2 domain than to the MAGI1 PDZ1 domain, the difference here is around 12 kJ mol^{-1}. We suspect that here also the absence of the lysine at position 72 (78 in MAGI1 PDZ1) reduces the binding free energy of the phosphorylated peptide to the MAGI1 PDZ1 domain. By forming a positively charged surrounding, ARG79 and LYS72 seem to play a very important role for the binding of phosphorylated peptides and likely peptides in general that contain a negative charge at position -2 from the C terminus. This is also seen in the plain MD simulations since the glutamate and phosphate repeatedly interact with ARG79 and LYS72 in the various simulations.

The results from the alchemical simulations and the binding patterns in the plain MD simulations also suggest that glutamate is not a good mimic for phosphoserine in such binding studies, as it was already stated before, because the phosphate group has the possibility to engage in stronger and more stable hydrogen bonds and salt bridges [66, 72, 156].

A question that remains open is why in experiments the phosphorylated EQVpSAV peptide binds weaker to the hPTP1E PDZ2 domain than the unphosphorylated EQVSAV peptide even though the phosphate group forms very stable hydrogen bonds with positively charged residues in the binding pocket. In our simulations with MAGI1 PDZ1 and hPTP1E PDZ2 R79A we also observed that the phosphorylated peptide binds weaker to the protein than the unphosphorylated one. The phosphorylated peptide binds even weaker when there are no positive residues surrounding the phosphate group (like in MAGI1 PDZ1). This makes perfect sense because the phosphate group can then not form strong salt bridges and freely interacts with the surrounding water. It seems like the phosphorylated peptide has a better coordination in the free (completely solvated) state as in the bound state. Maybe the interaction between the serine of the unphosphorylated EQVSAV peptide and the histidine in the binding pocket plays a role too because serine pulls the peptide further into the binding pocket but this was only observed for the hPTP1E PDZ2 domain and not for the MAGI1 PDZ1 domain.

4.5 Acknowledgement

This project was supported by Deutsche Forschungsgemeinschaft grant He3875/14-1.

Molecular graphics and analyses were performed with UCSF Chimera [343], developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

Special thanks go to V. Gapsys for providing hybrid structures for phosphorylated serines for the CHARMM36m force field to be used in the PMX [267, 346] program.

4.6 Supporting Information

The supporting information is shown in the appendix of this thesis in chapter B.

Chapter 5

How to study competitive binding computationally

5.1 Introduction

Within biological systems, a protein and its ligand are never alone, but are always influenced by external sources, such as other macromolecules or ligands. Multiple molecules usually compete to bind to the available proteins and when this competition is about a single binding pocket, it is called competitive binding.

In experimentl research, competitive binding assays are a common way to determine the binding strength of small molecule inhibitors and peptides to proteins. Thereby, the apparent value of the dissociation constant K_D of a ligand is measured when it is inhibited by another ligand. With the help of this information the dissociation constant K_I of the inhibitor can be determined. In drug design, the goal is to find an inhibitor with a dissociation constant that is similar or lower than the one of the original ligand.

Even though drug design is increasingly performed by aid of computational methods like virtual screening, docking and molecular dynamics, computational methods to examine the competitive binding of two ligands to the same receptor binding pocket are still missing to the best of our knowledge. Therefore, in this study we wanted to test a new computational methodology for competitive binding, namely parallel cascade selection molecular dynamics simulations (PaCS-MD) with a bound peptide and an inhibitor floating close to the binding pocket in combination with Markov state models (MSMs). We wanted to test if it is possible to obtain reasonable apparent dissociation constants or binding free energies for the bound ligand influenced by the inhibitor.

Due to their central role in various biological events and their conserved, well-defined binding mode [371] and the consequently growing interest in PDZ domains as potential drug targets [127, 372], PDZ (PSD-95/Discs-large/ZO-1) [117] domains are interesting systems to study competitive binding. They represent protein modules that are highly abundant in human cells and play a role in various protein-protein interactions, e. g. they are involved in signal transduction [120, 122–126]. Various experimental studies of competitive binding to PDZ domains exist [165, 175, 373–377] and different types of inhibitors are constantly developed [371].

The system of choice to study competitive binding in this project was the PDZ domain of the *mus musculus* (mouse) Dishevelled1 protein, further called mDvl1 PDZ binding a peptide with sequence SGSLKLMTTV peptide inhibited by the small-molecule inhibitor Sulindac [11] (SUZ). An experimental characterization has been previously provided by Lee et al. [175] and Zhang et al. [376]. In the former, they showed that Sulindac binds specifically to the mDvl1 PDZ domain. Using a competitive binding assay they observed that Sulindac inhibits the binding of a C-terminal peptide of the Dapper protein with the sequence $\text{SGSLKLMTTV}_{\text{COOH}}$. The inhibition constant of SUZ was in the same order of magnitude as the dissociation constant of the peptide. Sulindac is a commonly used non-steroidal anti-inflammatory drug that was demonstrated to help in prevention of colon cancer [378–383]. Additionally, Sulindac was shown to play a role in suppression of canonical β-catenin-related Wnt signaling in breast cancer and lung cancer cell lines [175, 384]. Dishevelled PDZ domains are involved in the transfer of Wnt signals from the membrane-bound receptor Frizzled to other cells [165] and since a nonnormal signaling of Wnt was connected to tumorigenesis, the Dishevelled PDZ domains are studied as targets for inhibition of Wnt signaling [375].

5.2 Materials and Methods

5.2.1 Structure preparation

For the purpose of this study we wanted to analyze the binding of the SGSLKLMTTV peptide and the small molecule inhibitor SUZ to the mDvl1 PDZ domain. A docking structure existed for SUZ binding the mDvl1 PDZ domain ((Protein Data Bank (PDB) [6] entry: 2KAW [175]) but at the time of starting this project there was no crystal structure available for the SGSLKLMTTV peptide binding to the mDvl1 PDZ domain. Consequently, we used another X-ray diffraction structure (PDB entry: 1L6O [385]), where the Xenopus Dishevelled PDZ domain binds a SLKLMTTV peptide from Dapper, to obtain a structure for the mDvl1 domain with the peptide. In doing so, we superimposed the PDZ domains (RMSD: 0.483 Å) and created a new structure of the mDvl1 PDZ domain binding the SLKLMTTV peptide to be used as our main study system. The binding of the SGSLKLMTTV peptide and the SLKLMTTV peptide should be very similar since protein binding to PDZ domains usually involves only a few C-terminal residues of the protein. All structures were taken from the Protein Data Bank (PDB) (www.rcsb.org) [6, 7].

Optimization of small molecule inhibitor SUZ

For the small molecule inhibitor SUZ we used the .mol2-file stored at PDB entry 2KAW [175] and used the CHARMM General Force Field (CGenFF) [1, 2] program [386, 387] to parametrize the ligand and obtain the CHARMM [3] files for this specific ligand. We then inserted this files into the FFParam program [388] to optimize the ligand parameters. This program helps performing molecular mechanics calculations using OpenMM [389] and quantum mechanical calculation using Psi4 [390]. After performing both types of simulations for the study system, the software is able to compare the quantum mechanical and molecular mechanical results and helps in optimizing the parameters so that the molecular mechanical results best resemble the quantum mechanical ones. First, geometry optimization was performed. We used the second-order Møller-Plesset perturbation theory [5] (MP2) with the 6-31G* basis set [391], as recommended in the FFParam documentation, to obtain the quantum mechanical structural properties. Afterwards, the electrostatic parameters, e. g. charges, were fitted. To do so, the quantum mechanical calculations for the water interactions were performed using the Hartree-Fock theory (HF) with the 6-31G* basis set, as recommended in the FFParam documentation and the molecular mechanics results were compared to them. In a last step, bonded parameters were optimized using the MP2 level of theory with the 6-31G* basis set again. The resulting CHARMM files were then transformed into GROMACS [4] format using a script provided by the authors of CGenFF on their website [392].

5.2.2 Molecular dynamics (MD) simulation protocol

The plain molecular dynamics simulations and the equilibration of the systems was performed similarly to our previous projects [32] (also see chapters 3 and 4) but is described here for completeness. We performed all our MD simulations with the GROMACS [4] 2018.8 software package and the CHARMM36m [192] force field for proteins and peptides. In order to represent the water molecules, we made use of the TIP3P [224] model modified for the CHARMM force field [190] and used the internal GROMACS tool pdb2gmx in order to generate hydrogen atoms. Moreover, a time step of 2 fs was applied in our simulations and periodic boundary conditions [201] were used. Additionally, we computed the long-range electrostatic interactions with the particle-mesh-Ewald summation method [8], whereby 12 Å were used as the non-bonded interaction cutoff. Furthermore, we constrained the hydrogen bonds using the LINCS [347] algorithm and applied long range dispersion corrections for energy and pressure. In order to simulate physiological conditions we included a concentration of $0.15\,\mathrm{mol\,L^{-1}}$ NaCl in all simulation boxes.

The minimization of our simulation systems was performed in two steps. First, we minimized the unsolvated molecular system for a total of 50000 steps using the steepest descent algorithm with an initial step size of 0.01 nm. Minimization was stopped when a value of convergence for the maximal force of $5\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$ was reached. Second, solvation and ionization of the system was performed and another round of minimization was carried out where the positions of the heavy atoms of protein and peptide were kept rigid via position restraints. Here again, the steepest descent algorithm was used. The total number of steps was 100000 and the initial step size 0.01 nm. We stopped the minimization when a convergence value of $500\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$ was reached.

After minimization we applied the velocity-rescaling thermostat [239] to our systems using a coupling time constant of 0.1 ps and separate temperature baths for solute and solvent. Thermalization was performed at 100 K, 200 K, and at the final temperature of 283 K, respectively, for 500 ps each while the position restraints for the proteins and peptides were still intact. We then continued the equilibration for 500 ps in the NPT ensemble. For this reason we added a Berendsen barostat [238] to the systems while the position restraints of the solutes were kept. Here, we used a barostat time constant of 2 ps, a reference pressure of 1 bar, and an isothermal compressibility of $4.6 \cdot 10^{-5}\mathrm{bar}^{-1}$. For the last equilibration steps, namely the release of the restraints on the solute in three steps, from 1000 via 100, and 10 to $0\,\mathrm{kJ/mol/nm^2}$, we changed the Berendsen barostat [238] to a Parrinello-Rahman extended-ensemble pressure coupling [250, 348]. This was needed because even though the former is great for box scaling in the start of simulations it does not reproduce a correct thermodynamic ensemble.

For the parallel cascade selection molecular dynamics (PaCS-MD) simulations of the small molecule inhibitor SUZ binding the mDvl1 PDZ domain and for the SLKLMTTV peptide binding the mDvl1 PDZ domain we created triclinic boxes with a size of around 8.6 nm x 8.6 nm x 14.1 nm. Here, the size in z-direction was larger because we aligned the binding pocket along this direction so the dissociation of the protein and peptide could happen mostly along this direction. For the PaCS-MD simulation for the SLKLMTTV peptide binding the mDvl1 PDZ domain where the small molecule inhibitor SUZ was floating around close to the binding pocket (i. e. the simulations to study competitive binding) the box size was approx. 9 nm x 9 nm x 14.1 nm.

Parallel Cascade Selection Molecular Dynamics (PaCS-MD)

Parallel cascade selection molecular dynamics (PaCS-MD) simulations were used in order to obtain absolute binding free energies for the SLKLMTTV peptide binding to the mDvl1 PDZ domain and for the small molecule inhibitor SUZ binding to the same domain. Furthermore, PaCS-MD simulations were also used to study the unbinding of the SLKLMTTV peptide when the inhibitor SUZ floats around nearby the binding pocket. The simulations were performed similar to the ones by Tran and Kitao [27] and our previous project (see chapter 4). To separate the protein and the ligand, i. e. the peptide or the small molecule inhibitor, during the PaCS-MD simulations the center-of-mass (COM) distance was used as the reaction coordinate.

In order to keep the protein from moving and rotating during our simulations we kept restraints on the C_α atoms of residues 2-7 and 84-89 with a strength of $100\,\mathrm{kJ/mol/nm^2}$.

We performed 30 independent runs (called trials) of PaCS-MD simulations for the unbinding of the SLKLMTTV peptide from the mDvl1 PDZ domain and another 30 trials for the unbinding of the small molecule inhibitor SUZ. Every 1 ns from the last 30 ns of an initial plain MD simulation with 100 ns length we took a snapshot to obtain starting structures for the trials. To start the PaCS-MD procedure we started 1 ns MD simulations (cycle 0) from each initial structure. Afterwards, we started 30 replicas from each of the 30 snapshots in the 1 ns MD simulations with the highest inter-COM distance between the protein and the peptide. We then simulated each replica for 100 ps in every cycle while recording snapshots every 0.5 ps. When the inter-COM distance between protein and peptide was greater than $d = 6\,\mathrm{nm}$ for all of the 30 best snapshots of the previous cycle in a trial, we stopped the simulations of this specific trial. For the PaCS-MD simulations of the SLKLMTTV peptide binding the mDvl1 PDZ domain where the small molecule inhibitor SUZ is floating around close to the binding pocket we performed many trials and analyzed different numbers of them to see how this affected the results.

To analyze the trajectories of the PaCS-MD simulations, we used pyEMMA [367] for the estimation of reversible discrete Markov state models (MSMs). We used the relative three-dimensional center-of-mass position (3D-COM) of the peptide with respect to the protein as a feature for the creation of MSMs as already done by Tran and Kitao [27]. Furthermore, we incorporated all of the trajectories from all PaCS-MD trials into the MSMs. Howeverm we removed all snapshots where the inter-COM distances were larger than 5 nm, because it is hard to perform a complete phase-space sampling in regions far away from the binding pocket. At 5 nm the protein and the peptide are completely separated and this is thus no issue for the analysis. In our analysis we had a total analyzed simulation time of 4.62 µs for mDvl1 PDZ with the SLKLMTTV peptide and of 1.73 µs for mDvl1 PDZ with the small molecule inhibitor SUZ.

We estimated unvalidated Markov models for a number of cluster centers using the Kmeans clustering [368] algorithm with repeated rounds of discretization to obtain a reasonable number of cluster centers or microstates and to account for the stochastic nature of the clustering algorithm. For examination of the clustering we used the VAMP-2 score [331] (with cross validation) to find an optimal value of cluster centers. In doing so, we used a MSM lag time of 20 steps for starting this analysis which was then adjusted later on to a reasonable lag time for the final MSM. We found 1000 clusters to be a good number of clusters for the binding of the SLKLMTTV peptide to the mDvl1 PDZ domain and 3000 clusters for the binding of the small molecule inhibitor SUZ to the mDvl1 PDZ domain because the VAMP-2 score reached a plateau at these values (see fig. C.1 and fig. C.2).

After these steps, the trajectory was transformed into a discretized trajectory along these clusters. We built MSMs for multiple lag times and plotted the resulting time scales for the ten slowest processes to find an optimal lag time to build the final MSMs. Usually one chooses the lag time where the time scales reach a converged plateau and in doing so we obtained a reasonable good final lag time of 10 ps for our MSMs for the binding of the SLKLMTTV peptide to the mDvl1 PDZ domain because there our results for the absolute binding free energy best resembled the experimental value. For the small molecule inhibitor SUZ binding the mDvl1 PDZ domain, we did not find a lag time where we obtained a binding free energy close to the result calculated from the experimental K_I value. Therefore, we chose the same lag time as for the SLKLMTTV peptide.

In order to calculate the binding free energy from the work performed along the reaction coordinate, we mapped the stationary distribution $\pi(\mathbf{r}_j)$ of each microstate j of the Markov state model based on the relative 3D-COM position of the peptide into a 1D-PMF by

$$W(d_i) = -\frac{1}{\beta} \ln \sum_{d_i - \delta d/2 \leq d(\mathbf{r}_j) < d_i + \delta d/2} \pi(\mathbf{r}_j) \text{ [27]} \tag{5.1}$$

where $\beta = 1/(k_\mathrm{B}T)$ is the inverse temperature, j is the position in 3D of microstate, $\mathbf{r}_j$ is the position of the microstate, and δd is the bin size. We chose the bin size as $\delta d = 0.05\,\mathrm{nm}$.

Competitive binding

When two ligands compete in binding to a receptor, in our case the mDvl1 PDZ domain, the dissociation constant K_D of the bound ligand reduces to the apparent dissociation constant $K_\mathrm{D}^\mathrm{app}$ as

$$K_\mathrm{D}^\mathrm{app} = K_\mathrm{D}\left(1 + \frac{[\mathrm{I}]}{K_\mathrm{I}}\right) \tag{5.2}$$

where [I] is the concentration of the inhibitor and K_I is the inhibition constant of the inhibitor.

To examine competitive binding and to obtain these apparent dissociation constants or binding free energies, we performed PaCS-MD simulations for the unbinding of the SLKLMTTV peptide from the mDvl1 PDZ domain where the small molecule inhibitor SUZ was floating around close to the binding pocket. We started multiple short plain MD simulations with different initial positions of the small molecule inhibitor SUZ scattered around the binding pocket to obtain a plethora of starting positions for the PaCS-MD simulations. From each short plain MD simulation we took nine starting positions. To keep the inhibitor close to the binding pocket we applied restraints on the center of mass (COM) of SUZ using the pull code in GROMACS. Those were one that kept the COM from moving to the other side of the protein away from the binding pocket by defining a minimum z-coordinate value in a flat-bottom potential (since the protein is aligned and fixed so the binding pocket opens up into z-direction) and one that kept the COM distance between SUZ and the protein below 3 nm.

From the initial positions extracted from the short plain MD simulations we started various PaCS-MD simulations. Beforehand, we released the restraints on the inhibitor to obtain an unbiased simulation system during PaCS-MD. We took different sets of initial positions and created MSMs using the PaCS-MD trajectories that were started from these positions to study how the initial positions of the inhibitor influenced the resulting apparent binding free energy of the peptide and the features of the MSM in general. For matter of comparison, we also performed one long plain MD simulation, extracted multiple initial positions for PaCS-MD from this, and created a MSM using only PaCS-MD trajectories that were started from initial positions from this single run. The total analyzed simulation times are given in table 5.1.

5.3 Results

5.3.1 Parallel Cascade Selection Molecular Dynamics of the SLKLMTTV peptide binding the mDvl1 PDZ domain

To obtain the absolute binding free energy for the binding of the SLKLMTTV peptide to the mDvl1 PDZ domain we performed PaCS-MD simulations to dissociate the peptide from the protein. In fig. C.1 the inter-COM distance between the SLKLMTTV peptide and the mDvl1 PDZ domain as a function of the number of cycles in the PaCS-MD trials is shown. There, always the largest inter-COM distance of each cycle is presented. We estimated the bound (inter-COM distance < 1.45 nm), partially-bound (1.45 nm < inter-COM distance < 2.1 nm) and unbound (2.1 nm < inter-COM distance) states from this plot. Overall, 52 ± 35 cycles and 78 ± 53 ns on average of PaCS-MD simulation were needed to completely dissociate the peptide from the protein. To analyze the PaCS-MD simulations MSMs were built with the 3D-COM position of the peptide as feature as was similarly done by Tran and Kitao [27] for dissociation PaCS-MD and as we already did in our previous study (see chapter 4).

The resulting absolute binding free energy for the binding of the SLKLMTTV peptide to the mDvl1 PDZ domain is

$$\Delta G^{\text{SLKLMTTV}}_{\text{bind,PaCS}} \approx -27.19\,\text{kJ}\,\text{mol}^{-1} \tag{5.3}$$

for a MSM lag time of 10 ps. This is nearly perfectly the experimental binding free energy $\Delta G^{\text{SGSLKLMTTV}}_{\text{bind,exp.}} = -27.75\,\text{kJ}\,\text{mol}^{-1}$ [175].

5.3.2 Parallel Cascade Selection Molecular Dynamics of the small molecule inhibitor SUZ binding the mDvl1 PDZ domain

To obtain the absolute binding free energy for the binding of the small molecule inhibitor SUZ to the mDvl1 PDZ domain we performed PaCS-MD simulations to dissociate the SUZ from the protein. In fig. C.2 the inter-COM distance between the small molecule inhibitor SUZ and the mDvl1 PDZ domain as a function of the number of cycles in the PaCS-MD trials is shown. There, always the largest inter-COM distance of each cycle is presented. We estimated the bound (inter-COM distance < 1.35 nm), partially-bound (1.35 nm < inter-COM distance < 2.1 nm) and unbound (2.1 nm < inter-COM distance) states from this plot. Overall, 20 ± 6 cycles and 29 ± 9 ns on average of PaCS-MD simulation were needed to completely dissociate the peptide from the protein. To analyze the PaCS-MD simulations MSMs were built

PaCS-MD simulations analyzed	Number of trajectories analyzed	Total simulation time in µs	Mean number of cycles	Mean length of trials in ns
0	34	5.98	33 ± 16	48 ± 23
1+2+3	27	4.66	58 ± 62	86 ± 93
1+2+4	27	5.61	70 ± 59	104 ± 88
1+3+4	27	4.46	56 ± 58	83 ± 87
2+3+4	27	4.20	52 ± 30	78 ± 45
1+2+3+4	36	6.31	59 ± 54	88 ± 81

TABLE 5.1: Number of analyzed trajectories, total simulation time in µs, mean number of cycles and mean length of trials in ns of trajectories from PaCS-MD for the creation of the MSMs. The standard deviation is displayed for the number of cycles and the length of trials.

with the 3D-COM position of SUZ as feature as was similarly done by Tran and Kitao [27] for dissociation PaCS-MD and as we already did in our previous study (see chapter 4).

The resulting absolute binding free energy for the binding of the small molecule inhibitor SUZ to the mDvl1 PDZ domain is

$$\Delta G^{\mathrm{SLKLMTTV}}_{\mathrm{bind,PaCS}} \approx -7.33\,\mathrm{kJ\,mol^{-1}} \tag{5.4}$$

for a MSM lag time of 10 ps. At none of the lag times the experimental binding free energy $\Delta G^{\mathrm{SUZ}}_{\mathrm{bind,exp.}} = -27.90\,\mathrm{kJ\,mol^{-1}}$ [175] could be reproduced wherefore we chose the same lag time as for the SLKLMTTV peptide.

5.3.3 Competitive binding of the small molecule inhibitor SUZ and the SLKLMTTV peptid to the mDvl1 PDZ domain by PaCS-MD

To analyze the influence of the initial position of the small molecule inhibitor SUZ on the features of the MSM created from the PaCS-MD trajectories for the SLKLMTTV peptide bound to the mDvl1 PDZ domain where SUZ was floating around close to the binding pocket, we combined the starting positions extracted from four short plain MD simulations into three sets. We always analyzed all trajectories that were started from initial positions extracted from a single plain MD simulations together but then analyzed these together with trajectories started form the initial positions from two or three other plain MD simulations. To keep track of the combinations, we labeled the plain MD simulations from 1 to 4. Since we also performed one long plain MD simulation and extracted initial positions from this one and analyzed them without combination of initial positions from other plain MD simulations we labeled this simulation with 0. In the following, the results from all the MSMs analyzing combinations of trajectories from PaCS-MD simulations will be shown in tables 5.1 and 5.2 for ease of comparison. To be comparable to each other and also to the MSM built from the PaCS-MD simulation for the SLKLMTTV peptide binding the mDvl1 PDZ domain without the small molecule inhibitor SUZ we used a lag time of 10 ps for all MSMs as it was found to reproduce the correct experimental absolute binding free energy for the SLKLMTTV peptide binding the mDvl1 PDZ domain.

All the resulting apparent binding free energies differed less than $4\,\mathrm{kJ\,mol^{-1}} \approx 1\,\mathrm{kcal\,mol^{-1}}$ from the apparent binding free energy $\Delta G^{\mathrm{app}}_{\mathrm{bind}} = k_{\mathrm{B}}T \ln K^{\mathrm{app}}_{\mathrm{D}} = -15.11\,\mathrm{kJ\,mol^{-1}}$

PaCS-MD simulations analyzed	**MSM lag time in** ps	**Apparent absolute binding free energy in** $\mathrm{kJ\,mol^{-1}}$
0	10	-11.46
1+2+3	10	-13.66
1+2+4	10	-14.20
1+3+4	10	-16.36
2+3+4	10	-13.83
1+2+3+4	10	-14.63

TABLE 5.2: Apparent absolute binding free energy in $\mathrm{kJ\,mol^{-1}}$ computed from MSMs at a lag time of 10 ps for the SLKLMTTV peptide binding the mDvl1 PDZ domain when the small molecule inhibitor SUZ is close to the binding pocket and competes in binding.

calculated from the experimental dissociation constant of the peptide, the inhibitor concentration in our simulations and the experimental inhibition constant of the inhibitor using eqn. (5.2). This deviation is a common size of error for absolute binding free energies obtained from computational methods. The result from the MSM built from the PaCS-MD trajectories started from initial positions that were all extracted from a single long plain MD trajectory (labeled 0) deviated the strongest from the expected value.

From fig. 5.1 and 5.2 it is clear, that the choice of initial positions of the small molecule inhibitor SUZ highly influences the 3D-COM positions of SUZ during the PaCS-MD trajectories and thus the features with which the MSMs are built. Therefore, different clusters will be found for different sets of trajectories and thus the MSMs and the resulting apparent binding free energies will differ.

5.4 Discussion

We realized that the choice of the initial position of the small molecule inhibitor close to the binding pocket has a strong impact on the length, i. e. the number of cycles, of a PaCS-MD simulation for a peptide bound to a protein domain when the inhibitor is trying to disturb this interaction. When the inhibitor is positioned directly on top of the binding pocket, the PaCS-MD simulation takes much longer than with the inhbitor being on the side because the inhibitor really blocks the peptide from moving out of the binding pocket. Since the simulation does not apply a pulling force or the like, it takes some time until the peptide pushes the inhibitor out of the way.

Our result for the absolute binding free energy of the small molecule inhibitor SUZ to the mDvl1 PDZ domain is far from reproducing the experimental result. Maybe the parametrization of SUZ was not ideal even though we optimized the parameters. Probably the initial position of SUZ obtained from PDB entry 2KAW was not optimal since it was obtained using docking [175].

Thus, it is interesting that the resulting apparent binding free energies from the PaCS-MD simulations where the peptide was bound and the small molecule inhibitor SUZ floated close to the binding pocket are well reproducing the apparent binding free energy that can be calculated from the experimental dissociation constant of the peptide, the experimental inhibition constant of SUZ and the concentration of SUZ in our simulations. We cannot be sure that this is just a coincidence

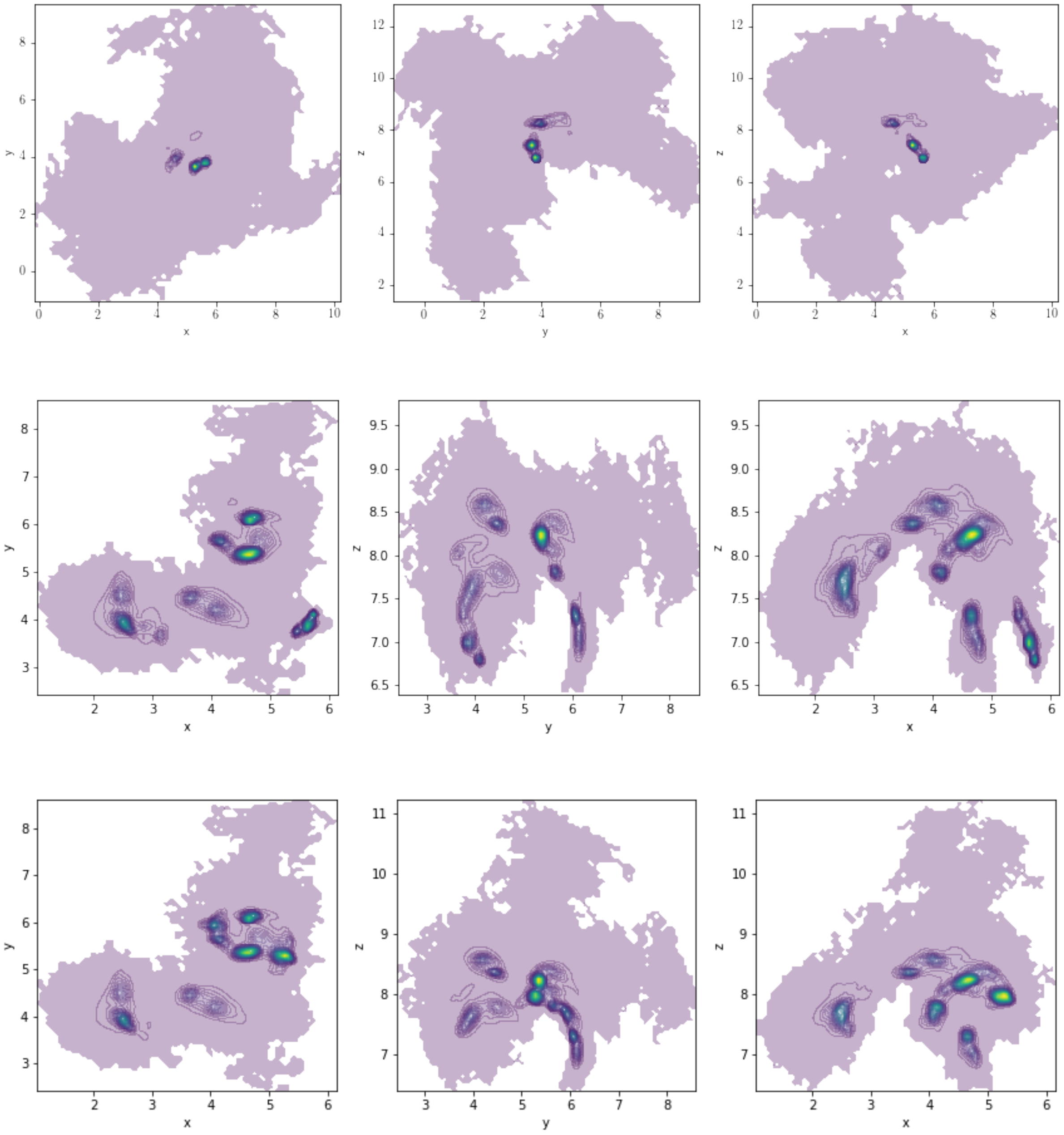


FIGURE 5.1: Densities of the 3D-COM inhibitor positions of all trajectories used to build a MSM. (A) Trajectories started from initial positions of the plain MD simulation labeled 0. (B) Trajectories started from initial positions of the plain MD simulations labeled 1, 2 and 3. (C) Trajectories started from initial positions of the plain MD simulations labeled 1, 2 and 4.

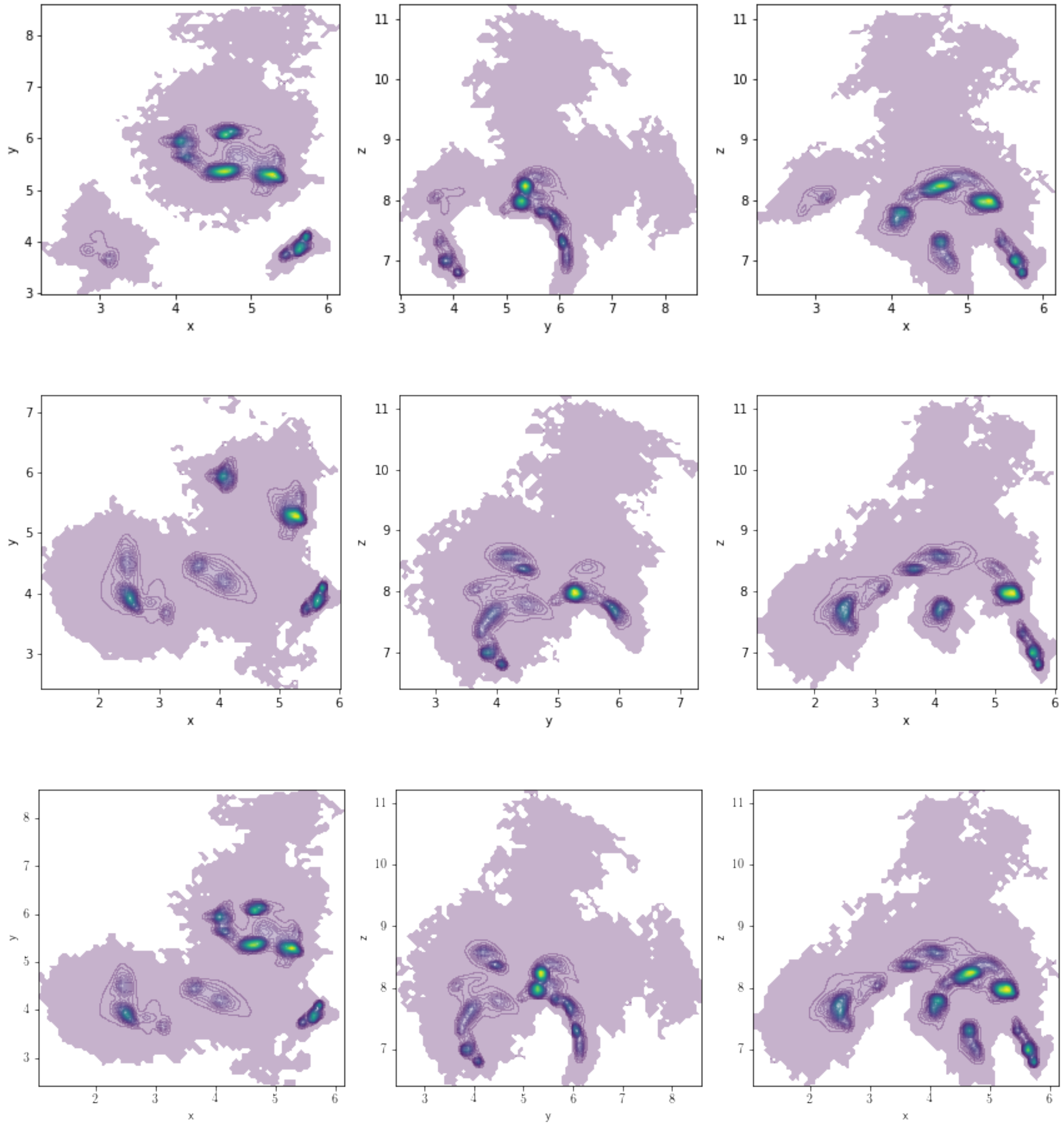


FIGURE 5.2: Densities of the 3D-COM inhibitor positions of all trajectories used to build a MSM. (A) Trajectories started from initial positions of the plain MD simulations labeled 1, 3 and 4. (B) Trajectories started from initial positions of the plain MD simulations labeled 2, 3 and 4. (C) Trajectories started from initial positions of the plain MD simulations labeled 1, 2, 3 and 4.

due to undersampling or other reasons. Because the results are so close to the calculated apparent binding free energy maybe the results for the competitive binding are correct and just the initial position of SUZ in the bound state for the PaCS-MD of SUZ binding the mDvl1 PDZ domain was non-optimal, therefore resulting in a too weak binding free energy for SUZ binding the mDvl1 PDZ domain. In this case, the parametrization could still be correct and thus the competitive binding worked well. Overall, it is clear from the inhibitor 3D-COM densities for the trajectories from the PaCS-MD simulations in fig. 5.1 that each MSM sampled different states of the system, i. e. different positions of the inhibitor relative to the binding pocket and peptide. This is a hint on a lack of enough sampling of all relevant states in the system. Here, it makes sense that the result from the MSM built from the PaCS-MD trajectories that were started from initial positions all extracted from a single long plain MD trajectory (labeled 0) deviated the strongest from the expected value, since here the sampling of 3D-COM positions of SUZ was the worst compared to the other MSMs. Additionally, the choice of the 3D-COM positions of peptide and inhibitor as features for the MSMs does not allow to differentiate between different binding states between peptide and inhibitor, peptide and protein, and inhibitor and protein that can occur during the course of the simulations.

Overall, this project is far from being completely finished but we obtained a way of studying the competitive binding of a peptide and a small molecule inhibitor to a protein including the calculation of the reduced apparent binding free energy of the peptide when the inhibitor is present close to the binding pocket. It is a functioning way of performing an unbinding simulation without external forces and thus allows to obtain various structures that could be intermediate states of the unbinding. Even though this is a positive result, further efforts have to be taken to get a deeper understanding of the influence of the initial inhibitor position on the computed binding free energies and the micro- and macrostates in a resulting MSM. Moreover, the influence of the number of different initial positions for sufficient sampling has to be further evaluated. Additionally, other features for the creation of the MSMs, like the C_α-atom positions, could be used to see how this influences the results. In the future, the same methodology could also be tested for other small molecule inhibitors, e. g. one studied by Shan et al. [375]. Furthermore, also the initial position of SUZ could be optimized by new docking simulations.

Chapter 6

Conclusion and Outlook

Overall, in this thesis the aim of a deeper understanding of protein-peptide interactions, specifically of 14-3-3 proteins and PDZ domains with C-terminal peptides of partner proteins, and of the influence of mutations in protein and peptide as well as phosphorylation of serine residues in the peptides on these interactions, was met. Computational studies of peptides binding to 14-3-3 proteins and PDZ domains are still scarce and even less computational studies exist that developed an understanding of the influence of phosphorylation on such binding.

6.1 14-3-3 proteins binding peptides

In our first two studies we performed plain MD simulations to determine mechanistic details of the binding of peptides to the respective protein or protein domain. In the first project [32] (see chapter 3), we gained new insight into the binding behaviour of unphosphorylated peptides to 14-3-3 domains, a binding that has not been observed in experiments until today, likely because the binding affinity of the unphosphorylated peptides is very low (at least the binding is too weak to determine dissociation constants for such peptides [104, 338, 339]). Our simulations clearly showed why complexes between phosphorylated peptides and 14-3-3 proteins are much more favorable than complexes with unphosphorylated peptides. Charged residues in the 14-3-3η domain, that we studied, formed strong interactions with the phosphate group and the phosphoserine side chain was better coordinated in the binding pocket compared to the unbound state while the opposite holds true for the serine side chain. Moreover, these interactions led to a strengthening of additional hydrogen bonds involving residues around the phosphorylated residue. We additionally obtained evidence for the common existence of arginine residues not further than three positions away from the phosphorylated residue inside the peptide chain in all three main binding motifs [104–106]. Usually, these interact with the phosphate group and lead to a better coordination in the bound state by closing the charged pocket formed by the 14-3-3 domain in that the phosphate group sits inside.

Furthermore, we learned that there are only very small differences between the binding of a single peptide (phosphorylated or not) to a 14-3-3 monomer or a 14-3-3 dimer and that it is challenging to reproduce or observe cooperative effects among the two binding sites of a 14-3-3 dimer that were previously suggested by experimentalists [104, 339, 341] (see chapter 3). We observed no specific differences in contacts between peptides and the proteins in a 14-3-3 dimer. The phosphorylation state of the peptide binding to the first 14-3-3 protein in the dimer had no influence on the binding of the second peptide binding to the second 14-3-3 protein in the dimer. Though, we saw an alteration of the main modes in a PCA analysis of 14-3-3 dimers based on the number of bound peptides (one or two) and their phosphorylation states.

To conclude, we realized that 1 µs plain MD simulations are likely too short to sample all relevant important dynamics of the 14-3-3 dimers since these are large systems with various conformational changes and movements and that it is generally very challenging to study complexes of these flexible peptides with 14-3-3 proteins. Thus, in the future it would be interesting to perform multiple long MD simulations of the 14-3-3 dimer with and without one or more peptides and to analyze the conformations using Markov state models (MSMs). Additionally, it would be insightful to compare the main conformations between systems with and without peptides to see what influence peptide binding has on the movements in the 14-3-3 dimer.

To complete the picture it would also make sense to additionally study complexes of 14-3-3 protein with motif 3 (RXXp(S/T)XX-COOH) [105] peptides in the future and to check whether the results differ from the ones in our study where we only studied motif 1 (R[S/F/Y/W]XpSXP) and motif 2 (RX[S/Y/FW/T/Q/A/D]Xp(S/T)X[P/L/M]) [104–106] peptides binding to 14-3-3 proteins.

6.2 PDZ domains binding peptides

In our second project, we aimed to study the binding of unphosphorylated and phosphorylated motif 1 peptides to PDZ domains and find out how this binding is influenced by mutations in the PDZ domains as well as by mutations in the peptide. We identified that peptides generally bind to PDZ domains due to backbone-backbone interactions with the binding-pocket beta strand and that the backbone of the residue at position -2 from the C terminus (often a phosphorylatable residue is found at this position) is involved in the most stable interactions. The binding is further stabilized by interactions of the hydrophobic C-terminal residue of the peptide with the surrounding residues of the PDZ domain.

Furthermore, we observed that specific residues in the PDZ domains have a strong influence on the binding of specifically phosphorylated peptides and other peptides that contain a negative charge at position -2 from the C terminus to these domains. These residues are ARG79 and LYS38 in the hPTP1E PDZ2 domain and equivalently GLN85 and LYS44 in the MAGI1 PDZ1 domain if mutated to arginine, as well as LYS72 in the hPTP1E PDZ2 domain.

In our simulations we did not focus on possible rearrangements in the PDZ domains due to binding since we always started from a bound state. Morra et al. showed, that strong rearrangements inside the PDZ can occur when a peptide binds to it. Therefore, it would be interesting to study the effects of these conformational changes on the binding and binding free energies in the future.

6.3 Binding free energies for peptides binding 14-3-3 proteins and PDZ domains and the influence of phosphorylation and the parametrization of the phosphate group

In the first two studies we also performed simulations to obtain binding free energies for the binding of the peptides to the respective proteins or protein domains. In the first project we used alchemical simulations to obtain binding free energy differences between the binding of phosphorylated and unphosphorylated peptides to

14-3-3 proteins. These revealed that there exist strong challenges to obtain good results for systems involving large charge changes, e. g. systems that involve doubly charged residues such as phosphoserine that should be alchemically transformed to a serine. Moreover, we observed deviations of the relative binding free energy differences from the experimental values. Thus, it would be interesting to perform absolute binding free energy simulation, like parallel cascade selection molecular dynamics (PaCS-MD) or umbrella sampling simulations, in order to see if the deviations exist for the phosphorylated and unphosphorylated peptides. If the deviations are different or lower for the unphosphorylated one this could be a hint, that the phosphate parameters in the CHARMM36m force field should maybe be updated to allow for calculation of correct binding free energies. Since there exist no experimental absolute binding free energy values for the unphosphorylated peptides (several publications state that the values are too small to be measured) it would be interesting to join forces with an experimental group and try to understand the binding of unphosphorylated peptides to 14-3-3 domains. In our study, we obtained bound states at least for 1 µs and saw some hydrogen bonds and contacts that existed nearly the full simulation time and therefore it is likely that the binding free energy of the unphosphorylated peptides is not 0 but just small.

In the second project (see chapter 4), we came across similar issues when calculating binding free energies as in the work on the binding of phosphorylated and unphosphorylated peptides to 14-3-3 domains [32] (see also chapter 3). The difference between our calculated alchemical free energy difference between the binding of a phosphorylated and an unphosphorylated peptide was around 9 kJ mol^{-1}. Here, that resulted in a better binding of the phosphorylated peptide compared to the unphosphorylated one even though the experimental result stated the opposite. We tested this using different setups for the alchemical transformation simulations, namely the “double system in a single box” setup and a setup, where we simulated the bound and free states seperately and applied box-size and electrostatic correction terms, but this did not strongly change the results. The reason could be the parametrization of the phosphate group in the CHARMM36m FF that could be optimized in the future to reproduce correct binding free energies for systems involving phosphate groups.

We additionally performed PaCS-MD simulations in our second project to obtain absolute binding free energy differences in addition to the relative binding free energy differences calculated using alchemical transformations. Here, it was possible to obtain an absolute binding free energy for the unphosphorylated peptide close to the experimental result but not for the phosphorylated peptide. Thus, we suspected that salt bridges between the phosphate group and charged residues in the binding pockets led to an overestimation of the binding strength of the phosphorylated peptide due to nonoptimal force field parameters for phosphoserine. We also tried to use glutamate as a mimic for phosphoserine in our simulations since common force fields are usually more reliable for natural amino acid side chains compared to the ones with post-translational modifications. We obtained good results for relative binding free energy differences from alchemical simulations using glutamate but our results from the plain MD simulations rather suggested that glutamate is not a good mimic for phosphoserine in such binding studies. This was also previously described [66, 72, 156].

It would be interesting to repeat both of the two studies with another force field that also includes phosphate parameters, e. g. the Amber ff14SB [393] or ff19SB [394] force fields, to see if the results strongly depend on the force field itself and if similar issues arise there. Furthermore, only since a few years force fields were optimized to also describe intrinsically disordered proteins like C-terminal peptide chains well. This could also have an influence on the binding of the peptides to the 14-3-3 domains in our simulations because the treatment of flexible chains can be different in different force fields.

In recent years a few experimental studies found evidence that specific binding motifs show a specificity to 14-3-3 and PDZ domains and that phosphorylation could be switching the interaction of a partner protein between these two [29–31]. Thus, in the future it would be very interesting to computationally study these specific binding motifs and systems to analyze how the phosphorylation changes the affinity of the peptide from PDZ to 14-3-3 or the opposite.

6.4 Competitive binding between a peptide and a small molecule inhibitor binding a PDZ domain

Our third study falls somewhat out of line with the other two since the focus is not mainly on the specific interactions between a peptide and a protein domain but rather how this interaction can be inhibited or how such a competitive binding can be studied computationally. As already mentioned in the discussion of chapter 5 (see sec. 5.4), this study is far from being finished. Though, we showed, that it is in general possible to study the competitive binding of a peptide and an inhibitor to the same binding pocket in a PDZ domain using PaCS-MD simulations, something that has, to our knowledge, not been done before. We saw a decrease in the apparent absolute binding free energy of the peptide and found a strong influence of the initial inhibitor positions on the results. Thus in the future, more sampling of inhibitor positions is needed. Furthermore, other features of the creation of MSMs could be used and more systems should be analyzed to further validate this approach.

6.5 Final conclusion

In summary, there are various challenges when performing molecular dynamics (MD) and free energy simulations but these are always evolving. They can help to obtain a deeper understanding of the interactions between proteins and their binding partners, which is not possible solely by experimental methods. In this thesis we identified a few possible challenges when studying the binding of peptides, especially phosphorylated peptides, to protein domains. Nevertheless, we have developed a deeper understanding of the binding of such peptides to PDZ domains and 14-3-3 proteins. Moreover, we showed that it is possible to study the competitive binding of peptides and inhibitors to PDZ domains using parallel cascade selection molecular dynamics (PaCS-MD) simulations. I am certain, that the field of protein-peptide simulations is still at the beginning and we will see strong developement in this area in the future due to increasingly more powerful hard- and software.

Appendix A

Supplementary Material for chapter 3

This supporting information is the preprint version of the supporting information for our manuscript "How phosphorylation of peptides affects their interaction with 14-3-3η domains" that has been accepted for publishing in PROTEINS: Structure, Function, and Bioinformatics [32].

A.1 Derivation of Correction Terms for Electrostatic Finite-Size Effects

From figure A.1 it is clear that:

$$\begin{aligned}\Delta\Delta G^{\mathrm{B-A}}_{\mathrm{bind}} &= \Delta G^{\mathrm{B}}_{\mathrm{bind}} - \Delta G^{\mathrm{A}}_{\mathrm{bind}} &\text{(A.1)}\\ &= \Delta G^{\mathrm{A\to B}}_{\mathrm{bound}} - \Delta G^{\mathrm{A\to B}}_{\mathrm{free}} &\text{(A.2)}\end{aligned}$$

and with corrections

$$\begin{aligned}\Delta\Delta G^{\mathrm{B-A}}_{\mathrm{bind,\,corr}} &= \Delta G^{\mathrm{B}}_{\mathrm{bind,\,corr}} - \Delta G^{\mathrm{A}}_{\mathrm{bind,\,corr}}\\ &= \Delta G^{\mathrm{A\to B}}_{\mathrm{bound,\,corrected}} - \Delta G^{\mathrm{A\to B}}_{\mathrm{free,\,corr}}\ . &\text{(A.3)}\end{aligned}$$

From figure A.2 it is clear that

$$\Delta G^{\mathrm{X}}_{\mathrm{bind}} = \Delta G^{\mathrm{X}}_{\mathrm{complex}} - \Delta G^{\mathrm{X}}_{\mathrm{solution}}\ , \qquad \text{(A.4)}$$

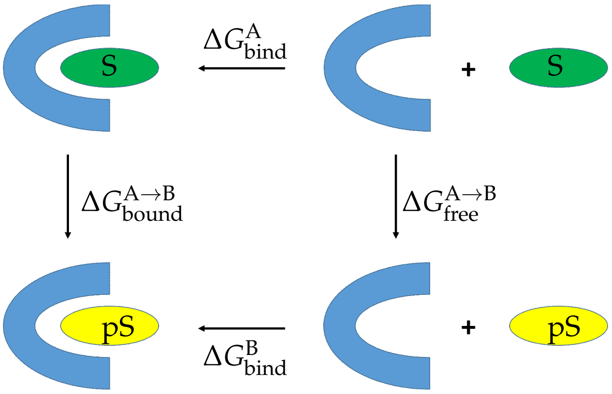
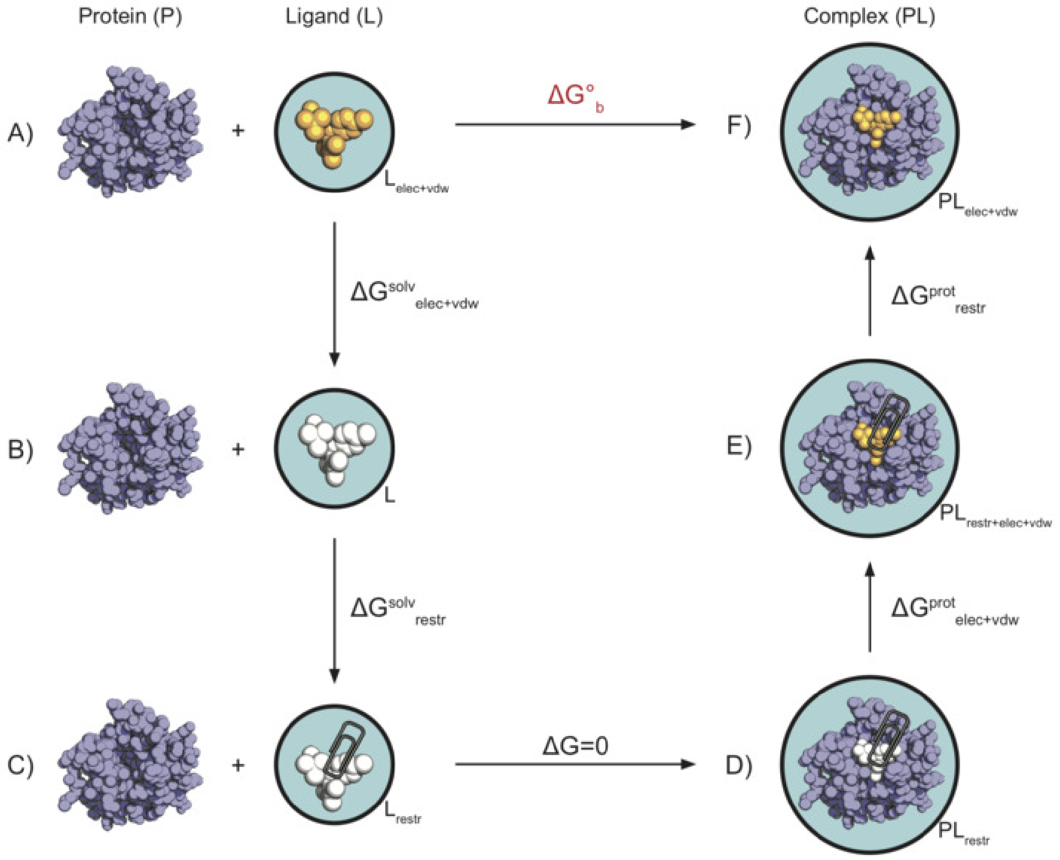


FIGURE A.1: Thermodynamic cycle for the difference in binding free energy between state A and B of a ligand bound to a receptor. The figure was taken from ref. [32].



Reproduced from Aldeghi *et al.*, Chem. Sci., 2016, **7**, 207-218. © The Royal Society of Chemistry

FIGURE A.2: Thermodynamic cycle for the absolute binding free energy. On the left is the free (solution) leg with total free energy $\Delta G_{\text{solution}}$ (considering the turning on of interactions, so from bottom to top) and on the right the bound (complex) leg $\Delta G_{\text{complex}}$ (also bottom to top) of the absolute binding simulation. Graph is taken from `http://alchemistry.org/wiki/Absolute_Binding_Free_Energy_-_Gromacs_2016` [395].

where X is a charge state of the ligand.
Using the correction terms we then obtain

$$\begin{aligned}
\Delta G^{\mathrm{X}}_{\text{bind, corr}} &= \Delta G^{\mathrm{X}}_{\text{complex, corr}} - \Delta G^{\mathrm{X}}_{\text{solution, corr}} \\
&= \Delta G^{\mathrm{X}}_{\text{complex}} + \Delta\Delta G^{\mathrm{X}}_{\text{NET\_USV, complex}} + \Delta\Delta G^{\mathrm{X}}_{\text{RIP, complex}} + \Delta\Delta G^{\mathrm{X}}_{\text{EMP, complex}} + \Delta\Delta G^{\mathrm{X}}_{\text{DSC, complex}} \\
&\quad - (\Delta G^{\mathrm{X}}_{\text{solution}} + \Delta\Delta G^{\mathrm{X}}_{\text{NET\_USV, solution}} + \Delta\Delta G^{\mathrm{X}}_{\text{RIP, solution}} \\
&\quad + \Delta\Delta G^{\mathrm{X}}_{\text{EMP, solution}} + \Delta\Delta G^{\mathrm{X}}_{\text{DSC, solution}}) \\
&= \underbrace{\Delta G^{\mathrm{X}}_{\text{complex}} - \Delta G^{\mathrm{X}}_{\text{solution}}}_{\Delta G^{\mathrm{X}}_{\text{bind}}} \\
&\quad + \Delta\Delta G^{\mathrm{X}}_{\text{NET\_USV, complex}} + \Delta\Delta G^{\mathrm{X}}_{\text{RIP, complex}} + \Delta\Delta G^{\mathrm{X}}_{\text{EMP, complex}} + \Delta\Delta G^{\mathrm{X}}_{\text{DSC, complex}} \\
&\quad - (\Delta\Delta G^{\mathrm{X}}_{\text{NET\_USV, solution}} + \Delta\Delta G^{\mathrm{X}}_{\text{RIP, solution}} + \Delta\Delta G^{\mathrm{X}}_{\text{EMP, solution}} + \Delta\Delta G^{\mathrm{X}}_{\text{DSC, solution}})
\end{aligned} \tag{A.5}$$

where

$$\Delta\Delta G_{\text{NET\_USV}}(L) = -\frac{\xi_{LS}}{8\pi\epsilon_0\epsilon_S}\left[(Q_P+Q_L)^2 - Q_P^2\right]\frac{1}{L}\,, \tag{A.6}$$

$$\Delta\Delta G_{\text{RIP}}(L) = \left[(I_P+I_L)(Q_P+Q_L) - I_P Q_P\right]\frac{1}{L^3}\,, \tag{A.7}$$

$$\Delta\Delta G_{\text{EMP}}(L) = -\frac{1}{8\pi\epsilon_0}\frac{16\pi^2}{45}\left(1-\frac{1}{\epsilon_S}\right)\times\left[(Q_P+Q_L)^2 - Q_P^2\right]\frac{R_L^5}{L^6} \tag{A.8}$$

and

$$\Delta\Delta G_{\text{DSC}}(N_S, L) = -\frac{\gamma_S Q_L}{6\epsilon_0}\frac{N_S}{L^3}\,. \tag{A.9}$$

The corrected relative binding free energy eqn. (A.3) is thus

$$\begin{aligned}
\Delta\Delta G^{\mathrm{B-A}}_{\text{bind, corr}} &= \Delta G^{\mathrm{B}}_{\text{bind, corr}} - \Delta G^{\mathrm{A}}_{\text{bind, corr}} \\
&= (\Delta G^{\mathrm{B}}_{\text{bind}} - \Delta G^{\mathrm{A}}_{\text{bind}}) \\
&\quad + \Delta\Delta G^{\mathrm{B}}_{\text{NET\_USV, complex}} + \Delta\Delta G^{\mathrm{B}}_{\text{RIP, complex}} + \Delta\Delta G^{\mathrm{B}}_{\text{EMP, complex}} + \Delta\Delta G^{\mathrm{B}}_{\text{DSC, complex}} \\
&\quad - \Delta\Delta G^{\mathrm{B}}_{\text{NET\_USV, solution}} - \Delta\Delta G^{\mathrm{B}}_{\text{RIP, solution}} - \Delta\Delta G^{\mathrm{B}}_{\text{EMP, solution}} - \Delta\Delta G^{\mathrm{B}}_{\text{DSC, solution}} \\
&\quad - (\Delta\Delta G^{\mathrm{A}}_{\text{NET\_USV, complex}} + \Delta\Delta G^{\mathrm{A}}_{\text{RIP, complex}} + \Delta\Delta G^{\mathrm{A}}_{\text{EMP, complex}} + \Delta\Delta G^{\mathrm{A}}_{\text{DSC, complex}} \\
&\quad - \Delta\Delta G^{\mathrm{A}}_{\text{NET\_USV, solution}} - \Delta\Delta G^{\mathrm{A}}_{\text{RIP, solution}} - \Delta\Delta G^{\mathrm{A}}_{\text{EMP, solution}} - \Delta\Delta G^{\mathrm{A}}_{\text{DSC, solution}})
\end{aligned} \tag{A.10}$$

The first term in eqn. (A.10) can be substituted by eqn. (A.2) yielding

$$\begin{aligned}
\Delta\Delta G^{\mathrm{B-A}}_{\text{bind, corr}} &= (\Delta G^{\mathrm{A}\to\mathrm{B}}_{\text{bound}} - \Delta G^{\mathrm{A}\to\mathrm{B}}_{\text{free}}) \\
&\quad + \Delta\Delta G^{\mathrm{B}}_{\text{NET\_USV, complex}} + \Delta\Delta G^{\mathrm{B}}_{\text{RIP, complex}} + \Delta\Delta G^{\mathrm{B}}_{\text{EMP, complex}} + \Delta\Delta G^{\mathrm{B}}_{\text{DSC, complex}} \\
&\quad - \Delta\Delta G^{\mathrm{B}}_{\text{NET\_USV, solution}} - \Delta\Delta G^{\mathrm{B}}_{\text{RIP, solution}} - \Delta\Delta G^{\mathrm{B}}_{\text{EMP, solution}} - \Delta\Delta G^{\mathrm{B}}_{\text{DSC, solution}} \\
&\quad - (\Delta\Delta G^{\mathrm{A}}_{\text{NET\_USV, complex}} + \Delta\Delta G^{\mathrm{A}}_{\text{RIP, complex}} + \Delta\Delta G^{\mathrm{A}}_{\text{EMP, complex}} + \Delta\Delta G^{\mathrm{A}}_{\text{DSC, complex}} \\
&\quad - \Delta\Delta G^{\mathrm{A}}_{\text{NET\_USV, solution}} - \Delta\Delta G^{\mathrm{A}}_{\text{RIP, solution}} - \Delta\Delta G^{\mathrm{A}}_{\text{EMP, solution}} - \Delta\Delta G^{\mathrm{A}}_{\text{DSC, solution}}) \\
&= \Delta G^{\mathrm{A}\to\mathrm{B}}_{\text{bound}} \\
&\quad + \Delta\Delta G^{\mathrm{B}}_{\text{NET\_USV, complex}} - \Delta\Delta G^{\mathrm{A}}_{\text{NET\_USV, complex}} + \Delta\Delta G^{\mathrm{B}}_{\text{RIP, complex}} - \Delta\Delta G^{\mathrm{A}}_{\text{RIP, complex}} \\
&\quad + \Delta\Delta G^{\mathrm{B}}_{\text{EMP, complex}} - \Delta\Delta G^{\mathrm{A}}_{\text{EMP, complex}} + \Delta\Delta G^{\mathrm{B}}_{\text{DSC, complex}} - \Delta\Delta G^{\mathrm{A}}_{\text{DSC, complex}} \\
&\quad - (\Delta G^{\mathrm{A}\to\mathrm{B}}_{\text{free}} \\
&\quad + \Delta\Delta G^{\mathrm{B}}_{\text{NET\_USV, solution}} - \Delta\Delta G^{\mathrm{A}}_{\text{NET\_USV, solution}} + \Delta\Delta G^{\mathrm{B}}_{\text{RIP, solution}} - \Delta\Delta G^{A}_{\text{RIP, solution}} \\
&\quad + \Delta\Delta G^{\mathrm{B}}_{\text{EMP, solution}} - \Delta\Delta G^{\mathrm{A}}_{\text{EMP, solution}} + \Delta\Delta G^{\mathrm{B}}_{\text{DSC, solution}} - \Delta\Delta G^{\mathrm{A}}_{\text{DSC, solution}}) \\
&= \Delta G^{\mathrm{A}\to\mathrm{B}}_{\text{bound, corr}} - \Delta G^{\mathrm{A}\to\mathrm{B}}_{\text{free, corr}} \qquad \text{(A.11)}
\end{aligned}$$

and thus we obtain the instructions for correcting the relative binding free energies. In the following we will try to combine the correction terms of the same kind starting with $\Delta\Delta G^{\mathrm{B}}_{\text{NET\_USV, state}} - \Delta\Delta G^{\mathrm{A}}_{\text{NET\_USV, state}}$ where "state" is either "bound" or "free" and L_{state} is the size of the simulation box in the specific state.
The correction for periodicity-induced net-charge interactions and undersolvation is thus

$$\begin{aligned}
&\Delta\Delta G^{\mathrm{A}\to\mathrm{B}}_{\text{NET\_USV, state}}(L_{\text{state}}) \\
&\quad = \Delta\Delta G^{\mathrm{B}}_{\text{NET\_USV, state}}(L_{\text{state}}) - \Delta\Delta G^{\mathrm{A}}_{\text{NET\_USV, state}}(L_{\text{state}}) \\
&\quad = -\frac{\xi_{LS}}{8\pi\epsilon_0\epsilon_S}\left[(Q_P + Q_L^{\mathrm{B}})^2 - Q_P^2\right]\frac{1}{L_{\text{state}}} - \left(-\frac{\xi_{LS}}{8\pi\epsilon_0\epsilon_S}\left[(Q_P + Q_L^{\mathrm{A}})^2 - Q_P^2\right]\frac{1}{L_{\text{state}}}\right) \\
&\quad = -\frac{\xi_{LS}}{8\pi\epsilon_0\epsilon_S}\left[(Q_P + Q_L^{\mathrm{B}})^2 - Q_P^2 - (Q_P + Q_L^{\mathrm{A}})^2 + Q_P^2\right]\frac{1}{L_{\text{state}}} \\
&\quad = -\frac{\xi_{LS}}{8\pi\epsilon_0\epsilon_S}\left[(Q_P + Q_L^{\mathrm{B}})^2 - (Q_P + Q_L^{\mathrm{A}})^2\right]\frac{1}{L_{\text{state}}} \\
&\quad = -\frac{\xi_{LS}}{8\pi\epsilon_0\epsilon_S}\left[Q_P^2 + (Q_L^{\mathrm{B}})^2 + 2Q_PQ_L^{\mathrm{B}} - (Q_P^2 + (Q_L^{\mathrm{A}})^2 + 2Q_PQ_L^{\mathrm{A}})\right]\frac{1}{L_{\text{state}}} \\
&\quad \underline{= -\frac{\xi_{LS}}{8\pi\epsilon_0\epsilon_S}\left[(Q_L^{\mathrm{B}})^2 - (Q_L^{\mathrm{A}})^2 + 2Q_P(Q_L^{\mathrm{B}} - Q_L^{\mathrm{A}})\right]\frac{1}{L_{\text{state}}}}\,. \qquad \text{(A.12)}
\end{aligned}$$

Continuing with the residual integrated potential correction we obtain

$$
\begin{aligned}
&\Delta\Delta G^{\mathrm{A}\to\mathrm{B}}_{\mathrm{RIP,\,state}}(L_{\mathrm{state}}) \\
&\quad = \Delta\Delta G^{\mathrm{B}}_{\mathrm{RIP,\,state}}(L_{\mathrm{state}}) - \Delta\Delta G^{\mathrm{A}}_{\mathrm{RIP,\,state}}(L_{\mathrm{state}}) \\
&\quad = \left[(I^{\mathrm{B}}_{P,\mathrm{state}} + I^{\mathrm{B}}_{L,\mathrm{state}})(Q_P + Q^{\mathrm{B}}_L) - I^{\mathrm{B}}_{P,\mathrm{state}} Q_P\right] \frac{1}{L^3_{\mathrm{state}}} \\
&\qquad - \left[(I^{\mathrm{A}}_{P,\mathrm{state}} + I^{\mathrm{A}}_{L,\mathrm{state}})(Q_P + Q^{\mathrm{A}}_L) - I^{\mathrm{A}}_{P,\mathrm{state}} Q_P\right] \frac{1}{L^3_{\mathrm{state}}} \\
&\quad = [(I^{\mathrm{B}}_{P,\mathrm{state}} + I^{\mathrm{B}}_{L,\mathrm{state}})(Q_P + Q^{\mathrm{B}}_L) - I^{\mathrm{B}}_{P,\mathrm{state}} Q_P \\
&\qquad - (I^{\mathrm{A}}_{P,\mathrm{state}} + I^{\mathrm{A}}_{L,\mathrm{state}})(Q_P + Q^{\mathrm{A}}_L) + I^{\mathrm{A}}_{P,\mathrm{state}} Q_P] \frac{1}{L^3_{\mathrm{state}}} \\
&\quad = [(I^{\mathrm{B}}_{P,\mathrm{state}} + I^{\mathrm{B}}_{L,\mathrm{state}}) Q_P + (I^{\mathrm{B}}_{P,\mathrm{state}} + I^{\mathrm{B}}_{L,\mathrm{state}}) Q^{\mathrm{B}}_L - I^{\mathrm{B}}_{P,\mathrm{state}} Q_P \\
&\qquad - (I^{\mathrm{A}}_{P,\mathrm{state}} + I^{\mathrm{A}}_{L,\mathrm{state}}) Q_P - (I^{\mathrm{A}}_{P,\mathrm{state}} + I^{\mathrm{A}}_{L,\mathrm{state}}) Q^{\mathrm{A}}_L + I^{\mathrm{A}}_{P,\mathrm{state}} Q_P] \frac{1}{L^3_{\mathrm{state}}} \\
&\quad = [I^{\mathrm{B}}_{P,\mathrm{state}} Q_P + I^{\mathrm{B}}_{L,\mathrm{state}} Q_P + (I^{\mathrm{B}}_{P,\mathrm{state}} + I^{\mathrm{B}}_{L,\mathrm{state}}) Q^{\mathrm{B}}_L - I^{\mathrm{B}}_{P,\mathrm{state}} Q_P \\
&\qquad - I^{\mathrm{A}}_{P,\mathrm{state}} Q_P - I^{\mathrm{A}}_{L,\mathrm{state}} Q_P - (I^{\mathrm{A}}_{P,\mathrm{state}} + I^{\mathrm{A}}_{L,\mathrm{state}}) Q^{\mathrm{A}}_L + I^{\mathrm{A}}_{P,\mathrm{state}} Q_P] \frac{1}{L^3_{\mathrm{state}}} \\
&\quad = \left[I^{\mathrm{B}}_{L,\mathrm{state}} Q_P + (I^{\mathrm{B}}_{P,\mathrm{state}} + I^{\mathrm{B}}_{L,\mathrm{state}}) Q^{\mathrm{B}}_L - I^{\mathrm{A}}_L Q_P - (I^{\mathrm{A}}_{P,\mathrm{state}} + I^{\mathrm{A}}_{L,\mathrm{state}}) Q^{\mathrm{A}}_L\right] \frac{1}{L^3_{\mathrm{state}}} \\
&\quad \underline{= \left[(I^{\mathrm{B}}_{P,\mathrm{state}} + I^{\mathrm{B}}_{L,\mathrm{state}}) Q^{\mathrm{B}}_L - (I^{\mathrm{A}}_{P,\mathrm{state}} + I^{\mathrm{A}}_{L,\mathrm{state}}) Q^{\mathrm{A}}_L + (I^{\mathrm{B}}_{L,\mathrm{state}} - I^{\mathrm{A}}_{L,\mathrm{state}}) Q_P\right] \frac{1}{L^3_{\mathrm{state}}}} \,.
\end{aligned}
\tag{A.13}
$$

Next the EMP correction gets calculated:

$$
\begin{aligned}
&\Delta\Delta G^{\mathrm{A}\to\mathrm{B}}_{\mathrm{EMP,\,state}}(L_{\mathrm{state}}) \\
&\quad = \Delta\Delta G^{\mathrm{B}}_{\mathrm{EMP,\,state}}(L_{\mathrm{state}}) - \Delta\Delta G^{\mathrm{A}}_{\mathrm{EMP,\,state}}(L_{\mathrm{state}}) \\
&\quad = -\frac{1}{8\pi\epsilon_0} \frac{16\pi^2}{45} \left(1 - \frac{1}{\epsilon_S}\right) \times \left[(Q_P + Q^{\mathrm{B}}_L)^2 - Q^2_P\right] \frac{(R^{\mathrm{B}}_{L,\mathrm{state}})^5}{L^6_{\mathrm{state}}} \\
&\qquad + \frac{1}{8\pi\epsilon_0} \frac{16\pi^2}{45} \left(1 - \frac{1}{\epsilon_S}\right) \times \left[(Q_P + Q^{\mathrm{A}}_L)^2 - Q^2_P\right] \frac{(R^{\mathrm{A}}_{L,\mathrm{state}})^5}{L^6_{\mathrm{state}}} \\
&\quad \underline{= -\frac{1}{8\pi\epsilon_0} \frac{16\pi^2}{45} \left(1 - \frac{1}{\epsilon_S}\right) \times \left(\left[(Q_P + Q^{\mathrm{B}}_L)^2 - Q^2_P\right] (R^{\mathrm{B}}_{L,\mathrm{state}})^5 - \left[(Q_P + Q^{\mathrm{A}}_L)^2 - Q^2_P\right] (R^{\mathrm{A}}_{L,\mathrm{state}})^5\right) \frac{1}{L^6_{\mathrm{state}}}}
\end{aligned}
\tag{A.14}
$$

with

$$
R^{\mathrm{X}}_{L,\mathrm{state}} = \left(\left[\frac{1}{8\pi\epsilon_0} \frac{4\pi}{3} \left(1 - \frac{1}{\epsilon_S}\right) Q^{\mathrm{X}}_L\right]^{-1} I^{\mathrm{X}}_{L,SLV,\mathrm{state}}\right)^{1/2} \,. \tag{A.15}
$$

FIGURE A.3: Hydrogen bonds present in more than 10 % of all frames (white = no hydrogen bond, blue = hydrogen bond) between a 14-3-3η monomer and the RLYHSLP (top) and RLYHpSLP (bottom) peptides in a monomer simulation. The figures were taken from the supporting information of ref. [32].

The last correction is the DSC correction

$$\begin{aligned}
&\Delta\Delta G^{\mathrm{A}\to\mathrm{B}}_{\mathrm{DSC,\,state}}(L_{\mathrm{state}})(N_S, L) \\
&\quad = \Delta\Delta G^{\mathrm{B}}_{\mathrm{DSC,\,state}}(L_{\mathrm{state}}) - \Delta\Delta G^{\mathrm{A}}_{\mathrm{DSC,\,state}}(L_{\mathrm{state}}) \\
&\quad = -\frac{\gamma_S Q_L^{\mathrm{B}}}{6\epsilon_0}\frac{N_{S,\mathrm{state}}}{L^3_{\mathrm{state}}} + \frac{\gamma_S Q_L^{\mathrm{A}}}{6\epsilon_0}\frac{N_{S,\mathrm{state}}}{L^3_{\mathrm{state}}} \\
&\quad = -\frac{\gamma_S}{6\epsilon_0}\frac{N_{S,\mathrm{state}}}{L^3_{\mathrm{state}}}\left(Q_L^{\mathrm{B}} - Q_L^{\mathrm{A}}\right) .
\end{aligned} \tag{A.16}$$

A.2 Contacts and Hydrogen Bonds

RLYHpSLP (39 pairs)

Residue pair	Occupancy
ARG57-NH2 – SP27-OE1	84.83
ARG57-NH1 – SP27-OE	74.88
ARG132-NH1 – SP27-OE1	73.32
ARG132-NH2 – SP27-OE2	69.31
TYR133-OH – SP27-OE2	66.65
LEU8-N – ASN178-OD1	66.15
LYS50-NZ – PRO9-OXT	65.64
ASN229-ND2 – HIS6-O	59.81
HIS6-N – ASN229-OD1	56.33
ASN178-ND2 – LEU8-O	53.36
LYS125-NZ – LEU8-O	39.93
LYS50-NZ – PRO9-O	39.59
TYR5-OH – GLU185-OE2	32.80
ARG132-NH1 – SP27-OE2	24.88
TYR5-OH – GLU185-OE1	24.30
ARG3-NH2 – GLU185-OE1	20.44
ARG57-NH1 – SP27-OE1	20.40
ARG3-NE – GLU185-OE2	19.20
LYS50-NZ – SP27-OE	19.16
ARG61-NH2 – SP27-OE2	18.12
ARG57-NH1 – SP27-OE2	18.06
LEU4-N – GLU185-OE1	16.87
LYS125-NZ – PRO9-O	16.80
ARG132-NH1 – LEU8-O	16.10
ARG3-NH2 – GLU185-OE2	15.94
LEU4-N – GLU185-OE2	15.68
ARG3-NE – GLU185-OE1	15.09
TYR5-N – GLU185-OE1	14.92
LYS125-NZ – PRO9-OXT	14.62
TYR5-N – GLU185-OE2	14.25
ARG3-NH2 – GLU136-OE2	13.19
ARG61-NH1 – SP27-OE2	12.30
ARG61-NH2 – SP27-OE	11.12
ARG3-NH2 – GLU136-OE1	11.07
TYR133-OH – PRO9-O	11.01
ARG132-NH2 – PRO9-O	10.73
ARG57-NH2 – SP27-OE	10.48
TYR133-OH – PRO9-OXT	10.44
ARG57-NH2 – SP27-OE2	10.21

RLYHSLP (20 pairs)

Residue pair	Occupancy
HIS6-N – ASN229-OD1	92.32
ASN229-ND2 – HIS6-O	88.74
LEU8-N – ASN178-OD1	48.33
ASN178-ND2 – LEU8-O	46.45
LYS50-NZ – PRO9-OXT	45.72
LYS125-NZ – PRO9-OXT	43.74
LYS125-NZ – PRO9-O	41.33
TYR5-OH – GLU185-OE2	35.99
TYR5-OH – GLU185-OE1	35.84
HIS6-NE2 – ASP228-OD2	33.71
LYS50-NZ – PRO9-O	32.40
HIS6-NE2 – ASP228-OD1	30.34
ARG3-NH2 – ASP228-OD1	17.24
ARG3-NH2 – ASP228-OD2	16.72
SER46-OG – PRO9-OXT	15.72
ARG3-NH1 – ASP228-OD1	15.54
ARG3-NH1 – ASP228-OD2	15.24
SER46-OG – PRO9-O	13.35
LYS125-NZ – LEU8-O	12.58
SER7-OG – ASN178-OD1	11.05

TABLE A.1: Hydrogen-bond occupancy for hydrogen bonds between 14-3-3η monomers and RLYHSLP/RLYHpSLP peptides which exist in more than 10% of all simulation frames. ARG3, LEU4, TYR5, HIS6, SER7/SP27, LEU8 and PRO9 are the residues of the RLYHSLP/RLYHpSLP peptides. The tables were taken from the supporting information of ref. [32].

RLYHpSLP

	H_2O + protein + peptide		H_2O + protein	
atom	**bound**	**free**	**bound**	**free**
OG	0.740 ± 0.002 (SD: 0.592)	0.681 ± 0.0020 (SD: 0.633)	0.313 ± 0.002 (SD: 0.464)	0.660 ± 0.002 (SD: 0.627)
OE1	3.619 ± 0.002 (SD: 0.595)	3.409 ± 0.003 (SD: 0.858)	2.381 ± 0.002 (SD: 0.639)	2.507 ± 0.004 (SD: 0.996)
OE2	3.324 ± 0.002 (SD: 0.477)	3.377 ± 0.003 (SD: 0.901)	3.158 ± 0.002 (SD: 0.651)	2.604 ± 0.004 (SD: 1.017)
OE	3.787 ± 0.002 (SD: 0.591)	3.404 ± 0.003 (SD: 0.878)	3.335 ± 0.003 (SD: 0.805)	2.589 ± 0.004 (SD: 0.995)
Total (no OG)	10.7287 ± 0.006 (SD: 0.965)	10.190 ± 0.009 (SD: 1.522)	8.872 ± 0.006 (SD: 1.216)	7.701 ± 0.010 (SD: 1.736)
Total	11.468 ± 0.007 (SD: 1.132)	10.870 ± 0.011 (SD: 1.649)	9.185 ± 0.008 (SD: 1.302)	8.361 ± 0.012 (SD: 1.846)

RLYHSLP

	H_2O + protein + peptide		H_2O + protein	
atom	**bound**	**free**	**bound**	**free**
OG	1.083 ± 0.002 (SD: 0.630)	1.175 ± 0.002 (SD: 0.617)	1.056 ± 0.002 (SD: 0.639)	1.128 ± 0.002 (SD: 0.631)

TABLE A.2: Mean number (count) of hydrogen bonds of each oxygen atom in the serine/phosphoserine side chains of the RLYHSLP/RLYHpSLP peptides in the bound (monomer simulation) and free state. Columns 2 and 3 labeled H_2O + protein + peptide (free: H_2O + peptide) lists the counts of hydrogen bonds between the oxygen atoms of (phospho)serine listed in the first column from the left and atoms from the surrounding water, the protein and the rest of the peptide (free: water and peptide). Columns 5 and 5 labeled H_2O + protein (free: H_2O) contain the counts of hydrogen bonds without the ones connecting the oxygen atoms to the peptide itself. The tables were taken from the supporting information of ref. [32].

RSRSTpSTP (24 pairs)

Residue pair	Occupancy
ARG132-NH2 – SP2259-OE2	100.00
TYR133-OH – SP2259-OE2	100.00
ARG132-NH1 – SP2259-OE1	99.90
ARG57-NH2 – SP2259-OE1	99.88
ASN178-OD1 – THR260-N	99.26
ARG57-NH1 – SP2259-OE	98.98
LYS50-NZ – SP2259-OE	93.13
ASN229-ND2 – THR258-O	90.50
TRP233-NE1 – SER257-OG	76.98
ASN178-OD1 – THR260-OG1	57.33
GLU185-OE2 – SER257-N	53.07
GLU185-OE2 – SER257-OG	51.35
ARG57-NH1 – SP2259-OE1	44.01
ASN229-OD1 – THR258-N	40.30
GLU185-OE1 – SER257-N	38.87
GLU185-OE1 – SER257-OG	38.29
LYS125-NZ – THR260-OG1	35.52
GLU185-OE2 – ARG254-NH2	32.64
GLU185-OE1 – ARG254-NH2	30.01
GLU185-OE1 – ARG254-NH1	28.30
ASN178-ND2 – THR260-OG1	28.18
GLU185-OE2 – ARG254-NH1	22.24
ARG132-NH1 – SP2259-OE2	20.77
LYS50-NZ – SP2259-OE2	13.02

RSRSTSTP (28 pairs)

Residue pair	Occupancy
ASN178-OD1 – THR260-N	88.67
ASN229-ND2 – THR258-O	82.60
ASN178-ND2 – THR260-O	66.81
ASN229-OD1 – THR258-N	64.96
TRP233-NE1 – SER257-OG	55.87
LYS125-NZ – THR260-O	39.98
ARG132-NH1 – SER259-OG	39.74
SER46-OG – PRO261-O	26.35
SER46-OG – PRO261-OXT	25.91
LYS125-NZ – PRO261-O	25.13
ASN178-OD1 – THR260-OG1	25.03
GLU185-OE2 – SER257-N	24.98
LYS125-NZ – THR260-OG1	22.14
LYS125-NZ – PRO261-OXT	22.00
GLU185-OE1 – SER257-N	21.79
GLU185-OE2 – SER257-OG	20.71
ARG132-NH2 – SER259-OG	19.64
LYS50-NZ – PRO261-O	19.47
LYS50-NZ – PRO261-OXT	18.84
ASN178-ND2 – THR260-OG1	18.21
GLU185-OE1 – SER257-OG	17.30
GLU185-OE1 – SER255-OG	13.82
GLU185-OE2 – ARG256-NH2	12.34
GLU185-OE2 – ARG256-N	11.97
GLU185-OE1 – ARG256-NE	10.83
GLU136-OE1 – ARG256-NH2	10.67
GLU136-OE2 – ARG256-NH2	10.64
GLU136-OE2 – ARG256-NH1	10.16

TABLE A.3: Hydrogen-bond occupancy for hydrogen bonds between 14-3-3η dimers and RSRSTSTP/RSRSTpSTP peptides (only one binding pocket occupied) which exist in more than 10% of all simulation frames. The first residue belongs to the protein and the second to the peptide. The tables were taken from the supporting information of ref. [32].

	H_2O + protein + peptide						
	atom	**RSRSTSTP**	**RSRSTpSTP**	**RSRSTSTP + HRYSTP**	**RSRSTpSTP + HRYSTP**	**RSRSTSTP + HRYpSTP**	**RSRSTpSTP + HRYpSTP**
RSRSTSTP/RSRSTpSTP	OG	1.124 ± 0.003 (SD: 0.642)	0.485 ± 0.002 (SD: 0.588)	1.087 ± 0.002 (SD: 0.632)	0.497 ± 0.002 (SD: 0.593)	1.127 ± 0.002 (SD: 0.628)	0.694 ± 0.002 (SD: 0.629)
	OE1	/	3.645 ± 0.002 (SD: 0.560)	/	3.822 ± 0.003 (SD: 0.668)	/	3.750 ± 0.002 (SD: 0.640)
	OE2	/	3.139 ± 0.002 (SD: 0.369)	/	3.631 ± 0.002 (SD: 0.600)	/	3.583 ± 0.002 (SD: 0.648)
	OE	/	3.927 ± 0.002 (SD: 0.654)	/	3.425 ± 0.002 (SD: 0.593)	/	3.719 ± 0.003 (SD: 0.654)
	Total (no OG)	/	10.711 ± 0.006 (SD: 0.961)	/	10.878 ± 0.006 (SD: 1.076)	/	11.052 ± 0.007 (SD: 1.121)
	Total	1.124 ± 0.003 (SD: 0.642)	11.195 ± 0.007 (SD: 1.126)	1.087 ± 0.002 (SD: 0.632)	11.375 ± 0.008 (SD: 1.228)	1.127 ± 0.002 (SD: 0.682)	11.746 ± 0.009 (SD: 1.286)
HRYSTP/HRYpSTP	OG	/	/	0.993 ± 0.002 (SD: 0.628)	1.100 ± 0.002 (SD: 0.636)	0.903 ± 0.002 (SD: 0.610)	0.992 ± 0.002 (SD: 0.648)
	OE1	/	/	/	/	3.494 ± 0.002 (SD: 0.568)	3.437 ± 0.002 (SD: 0.5514)
	OE2	/	/	/	/	3.321 ± 0.002 (SD: 0.523)	3.408 ± 0.002 (SD: 0.569)
	OE	/	/	/	/	3.516 ± 0.002 (SD: 0.547)	3.460 ± 0.002 (SD: 0.598)
	Total (no OG)	/	/	/	/	10.330 ± 0.006 (SD: 0.946)	10.428 ± 0.006 (SD: 1.008)
	Total	/	/	0.993 ± 0.002 (SD: 0.628)	1.100 ± 0.002 (SD: 0.636)	11.233 ± 0.008 (SD: 1.126)	11.389 ± 0.008 (SD: 1.225)

TABLE A.4: Mean number (count) of hydrogen bonds of each oxygen atom in the serine/phosphoserine side chains of the RSRSTSTP/RSRSTpSTP and HRYSTP/HRYpSTP peptides in the bound state of dimer simulations with two peptides. H_2O + protein + peptide are the hydrogen bonds between the oxygen atoms and atoms from the surrounding water, the protein and the rest of the peptide. The table was taken from the supporting information of ref. [32].

		H_2O + protein					
	atom	**RSRSTSTP**	**RSRSTpSTP**	**RSRSTSTP + HRYSTP**	**RSRSTpSTP + HRYSTP**	**RSRSTSTP + HRYpSTP**	**RSRSTpSTP + HRYpSTP**
RSRSTSTP/RSRSTpSTP	OG	1.111 ± 0.002 (SD: 0.640)	0.343 ± 0.002 (SD: 0.521)	1.078 ± 0.002 (SD: 0.633)	0.334 ± 0.002 (SD: 0.513)	1.120 ± 0.002 (SD: 0.632)	0.662 ± 0.002 (SD: 0.633)
	OE1	/	2.753 ± 0.003 (SD: 0.840)	/	3.270 ± 0.003 (SD: 0.865)	/	2.946 ± 0.004 (SD: 1.068)
	OE2	/	3.139 ± 0.002 (SD: 0.369)	/	2.514 ± 0.003 (SD: 0.768)	/	2.810 ± 0.004 (SD: 1.044)
	OE	/	3.676 ± 0.003 (SD: 0.677)	/	3.033 ± 0.003 (SD: 0.750)	/	2.691 ± 0.004 (SD: 0.999)
	Total (no OG)	/	9.568 ± 0.006 (SD: 1.140)	/	8.817 ± 0.008 (SD: 1.379)	/	8.447 ± 0.010 (SD: 1.798)
	Total	1.111 ± 0.002 (SD: 0.640)	9.911 ± 0.008 (SD: 1.254)	1.078 ± 0.002 (SD: 0.633)	9.151 ± 0.009 (SD: 1.747)	1.120 ± 0.002 (SD: 0.632)	9.109 ± 0.012 (SD: 1.906)
HRYSTP/HRYpSTP	OG	/	/	0.919 ± 0.002 (SD: 0.632)	1.025 ± 0.003 (SD: 0.660)	0.253 ± 0.002 (SD: 0.455)	0.290 ± 0.002 (SD: 0.437)
	OE1	/	/	/	/	3.035 ± 0.002 (SD: 0.544)	3.178 ± 0.003 (SD: 0.713)
	OE2	/	/	/	/	3.023 ± 0.002 (SD: 0.621)	3.227 ± 0.003 (SD: 0.655)
	OE	/	/	/	/	3.147 ± 0.002 (SD: 0.583)	2.738 ± 0.003 (SD: 0.737)
	Total (no OG)	/	/	/	/	9.205 ± 0.006 (SD: 1.011)	9.143 ± 0.007 (SD: 1.217)
	Total	/	/	0.919 ± 0.002 (SD: 0.632)	1.025 ± 0.003 (SD: 0.660)	9.458 ± 0.007 (SD: 1.108)	9.433 ± 0.008 (SD: 1.293)

TABLE A.5: Mean number (count) of hydrogen bonds of each oxygen atom in the serine/phosphoserine side chains of the RSRSTSTP/RSRSTpSTP and HRYSTP/HRYpSTP peptides in the bound state of dimer simulations with two peptides. H_2O + protein are the hydrogen bonds without the ones connecting the oxygen atoms to the peptide itself. The table was taken from the supporting information of ref. [32].

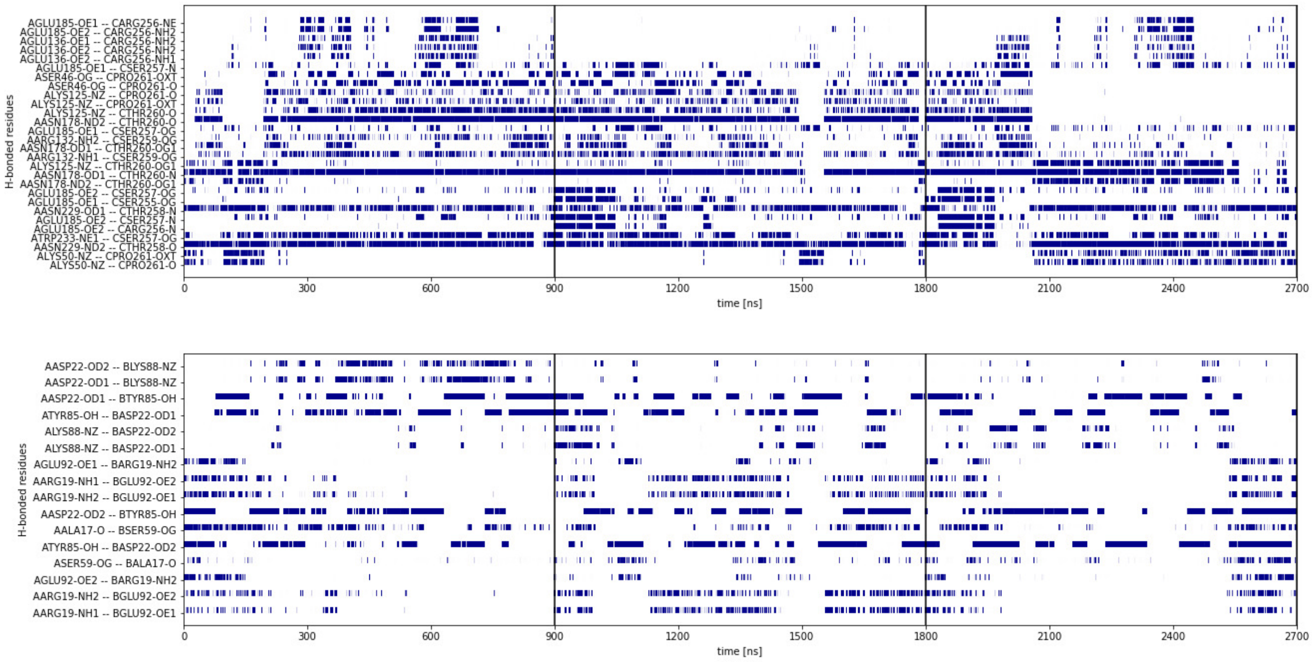


FIGURE A.4: Hydrogen bonds present in more than 10 % of all frames (white = no hydrogen bond, blue = hydrogen bond) between a 14-3-3η monomer and the RSRSTSTP (top) and RSRSTpSTP (bottom) peptides in a dimer simulation. The figures were taken from the supporting information of ref. [32].

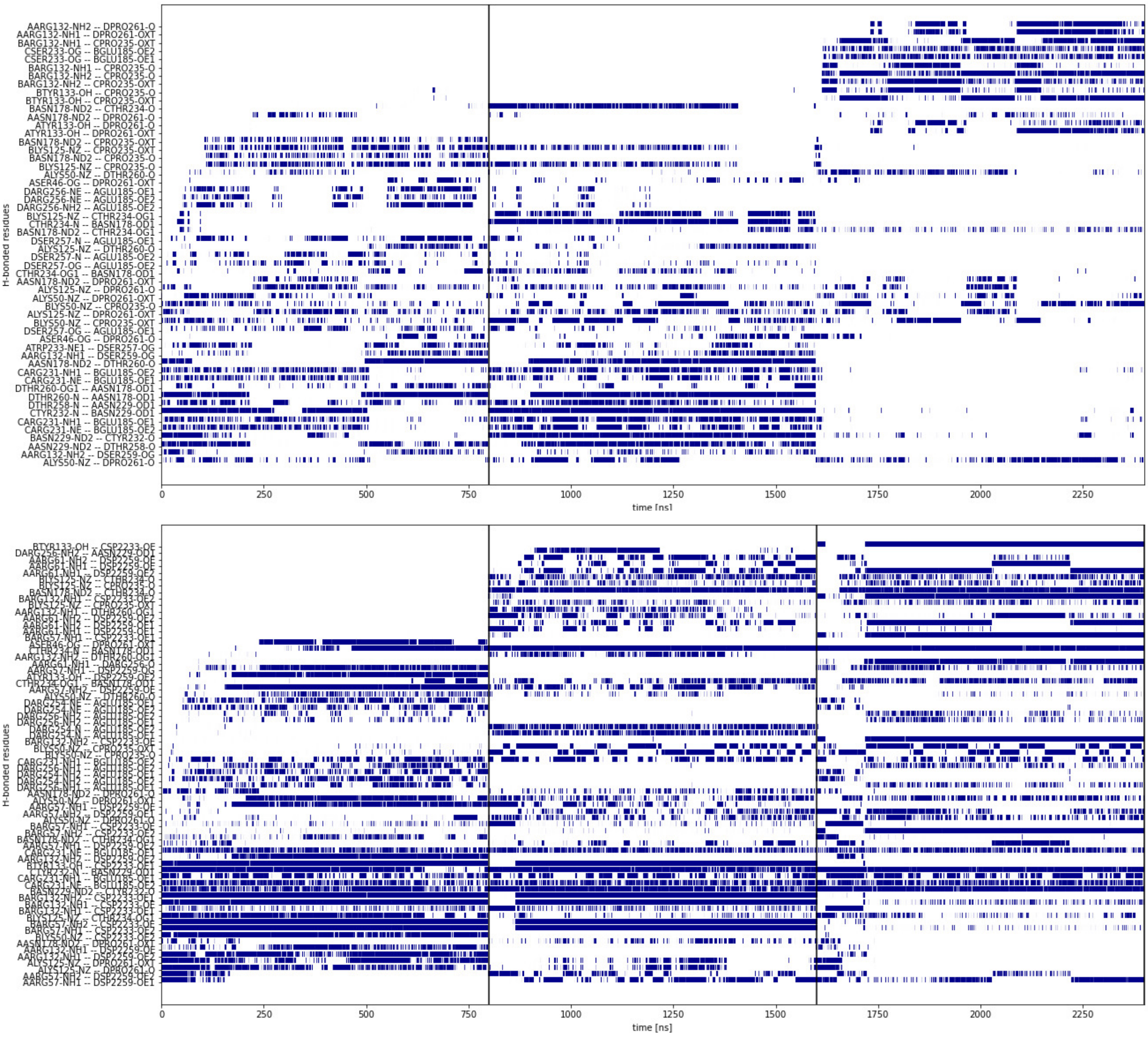


FIGURE A.5: Hydrogen bonds present in more than 10 % of all frames (white = no hydrogen bond, blue = hydrogen bond) between the monomers and the RSRSTSTP + HRYSTP (top) and RSRSTpSTP + HRYpSTP (bottom) peptides in a 14-3-3η dimer simulation. The figures were taken from the supporting information of ref. [32].

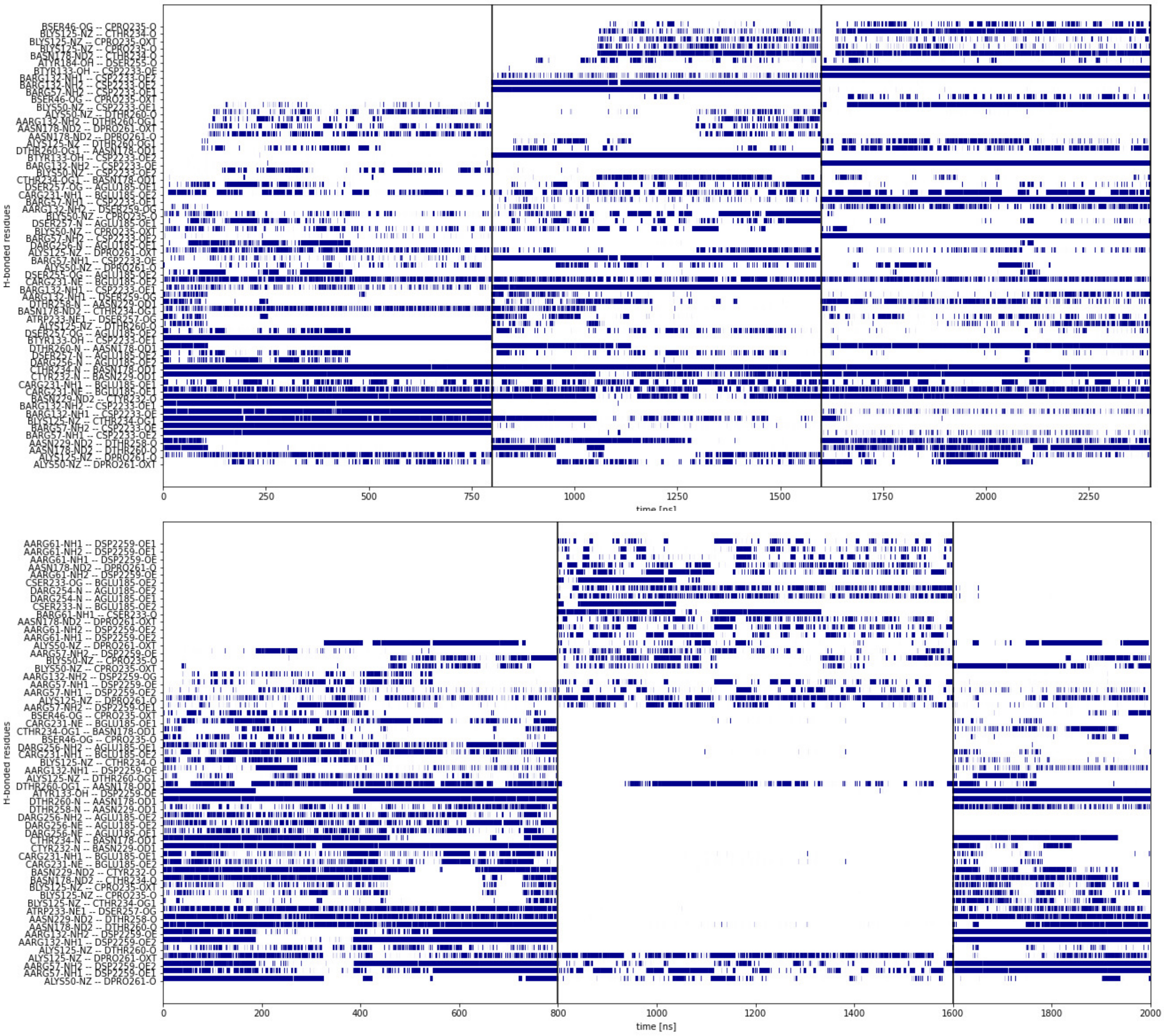


FIGURE A.6: Hydrogen bonds present in more than 10 % of all frames (white = no hydrogen bond, blue = hydrogen bond) between the monomers and the RSRSTSTP + HRYpSTP (top) and RSRSTpSTP + HRYSTP (bottom) peptides in a 14-3-3η dimer simulation. The figures were taken from the supporting information of ref. [32].

RSRSTpSTP (31 contacts)

Residue pair	Distance	Int. energy
AARG57 – CSP2259	1.69	-514.86
AARG132 – CSP2259	1.66	-486.69
ALYS50 – CSP2259	3.04	-300.36
ALYS50 – CPRO261	2.43	-233.27
ALYS125 – CPRO261	3.45	-227.94
AGLU185 – CARG256	3.81	-152.31
ATYR133 – CSP2259	1.7	-120.24
ALYS125 – CTHR260	2.49	-48.7
AASN178 – CTHR260	1.9	-43.36
ALYS50 – CTHR260	3.55	-36.44
AASN229 – CTHR258	2.1	-30.13
AASN178 – CSP2259	2.65	-25.64
AGLU185 – CSER257	3.22	-18.51
AVAL181 – CSP2259	2.58	-17.08
AASN229 – CSER257	2.85	-16.4
ATRP233 – CSER257	2.46	-15.11
ASER46 – CPRO261	4.0	-14.7
AGLY174 – CTHR260	2.47	-6.62
ALEU225 – CTHR258	2.44	-5.74
ALEU177 – CSP2259	2.79	-4.69
ALEU232 – CARG256	3.31	-4.32
AILE222 – CTHR260	2.67	-4.29
ALEU225 – CSP2259	3.9	-4.0
ATYR184 – CSER257	2.87	-3.56
AVAL181 – CSER257	2.68	-3.47
AVAL181 – CTHR258	3.16	-3.46
ALEU177 – CTHR260	2.27	-3.41
AVAL47 – CPRO261	3.86	-2.93
ALEU177 – CTHR258	2.77	-2.26
ALEU232 – CSER257	2.87	-0.91
ALEU225 – CTHR260	3.01	-0.65

RSRSTSTP (26 contacts)

Residue pair	Distance	Int. energy
ALYS125 – CPRO261	3.4	-229.36
AGLU185 – CARG256	3.33	-144.57
ALYS50 – CPRO261	2.23	-131.91
ALYS125 – CTHR260	2.53	-47.81
AGLU185 – CSER257	2.6	-41.36
ASER46 – CPRO261	3.32	-39.75
AASN178 – CTHR260	2.1	-37.94
AASN229 – CTHR258	2.14	-35.4
AASN178 – CSER259	2.75	-22.16
AARG132 – CSER259	2.77	-18.39
ATRP233 – CSER257	2.47	-15.21
AASN229 – CSER257	2.92	-14.84
ALEU225 – CTHR258	2.31	-7.23
AGLY174 – CTHR260	2.75	-4.96
ALEU232 – CARG256	3.57	-4.67
AVAL181 – CSER257	2.47	-4.48
AVAL181 – CTHR258	2.86	-4.26
AVAL47 – CPRO261	3.93	-4.1
AILE222 – CTHR260	2.74	-3.84
ALEU177 – CTHR260	2.35	-3.43
ALEU177 – CSER259	2.63	-3.12
ATYR184 – CSER257	2.98	-2.69
ALEU177 – CTHR258	2.81	-2.16
ALEU225 – CTHR260	3.04	-1.48
ALEU232 – CSER257	3.13	-0.82
AVAL181 – CSER259	2.76	0.02

TABLE A.6: Residue contacts between a 14-3-3η dimer and the RSRSTSTP or RSRSTpSTP peptides (only one peptide bound to the dimer) and respective mean interaction energies in kJ mol^{-1}. Here, all contacts were counted when the mean of the shortest atom distance (in Å) between two residues over all frames of three repeated simulations was lower than 4 Å. The first residue belongs to the protein and the second one to the peptide. SP2259 is the phosphorylated (doubly unprotonated) SER259. The tables were taken from the supporting information of ref. [32].

RSRSTSTP (8 contacts)

Residue pair	Distance	Int. energy
ALYS125 – DPRO261	3.05	-237.42
ALYS50 – DPRO261	2.37	-200.85
ALYS50 – DTHR260	3.4	-30.16
AASN178 – DTHR260	3.07	-25.34
AASN178 – DPRO261	3.52	-23.55
ALEU225 – DTHR258	3.47	-4.8
ALEU177 – DTHR260	3.68	-1.81
ALEU225 – DTHR260	3.93	-0.74

HRYSTP (15 contacts)

Residue pair	Distance	Int. energy
BLYS50 – CPRO235	2.31	-249.85
BLYS125 – CPRO235	3.2	-222.1
BGLU185 – CARG231	3.94	-175.58
BARG132 – CTHR234	3.78	-25.32
BASN229 – CTYR232	2.53	-23.04
BASN178 – CTHR234	2.4	-22.56
BLYS125 – CTHR234	3.84	-20.0
BASP228 – CTYR232	3.17	-16.06
BASN229 – CARG231	3.85	-15.63
BLEU225 – CTYR232	3.0	-8.74
BVAL181 – CTYR232	3.93	-3.05
BVAL181 – CTHR234	3.78	-2.04
BLEU177 – CTHR234	3.07	-2.02
BVAL181 – CSER233	3.79	-1.84
BLEU232 – CARG231	3.8	0.61

TABLE A.7: Residue contacts between a 14-3-3η dimer and RSRSTSTP and HRYSTP peptides (one in each binding pocket) and respective mean interaction energies in kJ mol^{-1}. Here, all contacts were counted when the mean of the shortest atom distance (in Å) between two residues over all frames of three repeated simulations was lower than 4 Å. The first residue belongs to the protein and the second one to the peptide. SP2259 is the phosphorylated (doubly unprotonated) SER259. The tables were taken from the supporting information of ref. [32].

HRYpSTP (33 contacts)

Residue pair	Distance	Int. energy
BARG57 – CSP2233	1.68	-520.4
BARG132 – CSP2233	1.67	-493.06
BLYS50 – CSP2233	3.06	-288.3
BGLU185 – CARG231	1.83	-281.08
BLYS50 – CPRO235	2.71	-226.62
BLYS125 – CPRO235	3.5	-221.65
BTYR133 – CSP2233	1.72	-119.71
BGLU136 – CARG231	4.0	-87.04
BLYS125 – CTHR234	2.22	-59.7
BLYS50 – CTHR234	3.35	-43.38
BASN178 – CTHR234	1.97	-40.1
BASN229 – CTYR232	1.98	-36.41
BASN178 – CSP2233	2.82	-28.04
BVAL181 – CSP2233	2.75	-19.4
BASP228 – CTYR232	2.75	-18.03
BASN229 – CARG231	2.86	-17.97
BSER46 – CPRO235	3.63	-17.89
BLEU225 – CTYR232	2.39	-14.84
BGLY174 – CTHR234	2.45	-6.15
BVAL181 – CARG231	2.53	-5.06
BVAL181 – CTYR232	2.81	-4.65
BVAL47 – CPRO235	3.49	-4.07
BLEU177 – CSP2233	2.84	-3.65
BILE222 – CTHR234	2.75	-3.41
BLEU177 – CTHR234	2.34	-2.65
BLEU177 – CTYR232	2.73	-2.14
BLEU225 – CTHR234	3.4	-0.19
BARG57 – CARG231	2.35	0.0
BLEU232 – CARG231	2.88	1.41
BTYR184 – CARG231	2.85	2.61
BLEU232 – CHIS230	3.66	2.63
BTRP233 – CARG231	3.06	4.25
BARG132 – CARG231	2.43	106.4

RSRSTpSTP (8 contacts)

Residue pair	Distance	Int. energy
AARG57 – DSP2259	1.73	-453.53
AARG61 – DSP2259	3.46	-331.85
ALYS50 – DPRO261	2.14	-285.48
AGLU185 – DARG254	3.4	-257.41
AARG132 – DSP2259	3.7	-244.83
ATYR133 – DSP2259	3.63	-51.46
ALYS50 – DTHR260	3.32	-42.37
AARG61 – DARG254	3.94	120.01

TABLE A.8: Residue contacts between a 14-3-3η dimer and RSRSTpSTP and HRYpSTP peptides (one in each binding pocket) and respective mean interaction energies in kJ mol^{-1}. Here, all contacts were counted when the mean of the shortest atom distance (in Å) between two residues over all frames of three repeated simulations was lower than 4 Å. The first residue belongs to the protein and the second one to the peptide. SP2259 is the phosphorylated (doubly unprotonated) SER259. The tables were taken from the supporting information of ref. [32].

RSRSTSTP (10 contacts)

Residue pair	Distance	Int. energy
ALYS125 – DPRO261	2.71	-272.79
ALYS50 – DPRO261	2.4	-199.6
ALYS50 – DTHR260	3.4	-31.97
AASN178 – DTHR260	2.55	-30.93
AASN178 – DPRO261	3.6	-27.94
ALYS125 – DTHR260	3.68	-25.84
AASN229 – DTHR258	3.86	-18.32
ALEU225 – DTHR258	3.44	-4.34
ALEU177 – DTHR260	2.91	-1.31
ALEU225 – DTHR260	3.31	-0.83

HRYpSTP (32 contacts)

Residue pair	Distance	Int. energy
BARG57 – CSP2233	1.69	-519.13
BARG132 – CSP2233	1.67	-487.69
BLYS50 – CSP2233	2.87	-344.51
BGLU185 – CARG231	1.83	-280.24
BLYS50 – CPRO235	2.34	-194.26
BLYS125 – CPRO235	3.77	-187.3
BTYR133 – CSP2233	1.71	-120.65
BLYS125 – CTHR234	2.01	-59.87
BASN178 – CTHR234	1.92	-36.01
BASN229 – CTYR232	1.96	-35.54
BLYS50 – CTHR234	3.47	-29.94
BSER46 – CPRO235	3.57	-25.4
BASN178 – CSP2233	2.7	-22.84
BASN229 – CARG231	2.85	-19.68
BVAL181 – CSP2233	2.55	-19.18
BASP228 – CTYR232	2.94	-17.11
BLEU225 – CTYR232	2.36	-14.45
BVAL181 – CARG231	2.48	-5.08
BLEU177 – CSP2233	2.67	-4.58
BVAL181 – CTYR232	2.75	-4.5
BGLY174 – CTHR234	2.51	-4.22
BVAL47 – CPRO235	3.8	-4.12
BILE222 – CTHR234	2.85	-3.16
BLEU177 – CTHR234	2.33	-2.91
BLEU177 – CTYR232	2.69	-2.54
BLEU225 – CTHR234	3.47	-0.51
BARG57 – CARG231	2.37	0.0
BLEU232 – CARG231	2.92	0.94
BLEU232 – CHIS230	3.56	2.81
BTYR184 – CARG231	2.87	2.97
BTRP233 – CARG231	3.11	4.17
BARG132 – CARG231	2.44	105.55

TABLE A.9: Residue contacts between a 14-3-3η dimer and RSRSTSTP and HRYpSTP peptides (one in each binding pocket) and respective mean interaction energies in kJ mol^{-1}. Here, all contacts were counted when the mean of the shortest atom distance (in Å) between two residues over all frames of three repeated simulations was lower than 4 Å. The first residue belongs to the protein and the second one to the peptide. SP2259 is the phosphorylated (doubly unprotonated) SER259. The tables were taken from the supporting information of ref. [32].

RSRSTpSTP (23 contacts)

Residue pair	Distance	Int. energy
AARG57 – DSP2259	1.71	-492.64
AARG132 – DSP2259	2.73	-331.95
ALYS125 – DPRO261	2.28	-313.91
ALYS50 – DPRO261	2.43	-241.73
ATYR133 – DSP2259	2.57	-83.19
AASN178 – DTHR260	2.03	-41.46
ALYS50 – DTHR260	3.59	-38.69
ALYS125 – DTHR260	3.43	-33.3
AASN178 – DPRO261	3.5	-25.21
AASN229 – DTHR258	3.28	-17.16
AVAL181 – DSP2259	3.64	-12.09
AASN229 – DSER257	3.65	-11.8
ALEU232 – DARG256	2.96	-6.07
AGLY174 – DTHR260	3.59	-5.65
ALEU225 – DTHR258	3.13	-3.76
AILE222 – DTHR260	3.78	-3.36
ALEU225 – DPRO261	3.79	-2.73
ALEU177 – DTHR260	2.36	-2.67
ALEU232 – DSER255	3.88	-1.93
ALEU232 – DSER257	3.51	-1.28
ALEU225 – DTHR260	2.9	-1.07
AVAL181 – DTHR260	3.85	-0.88

HRYSTP (1 contact)

Residue pair	Distance	Int. energy
BLYS50 – CPRO235	2.78	-190.59

TABLE A.10: Residue contacts between a 14-3-3η dimer and RSRSTpSTP and HRYSTP (no contacts found) peptides (one in each binding pocket) and respective mean interaction energies in kJ mol^{-1}. Here, all contacts were counted when the mean of the shortest atom distance (in Å) between two residues over all frames of three repeated simulations was lower than 4 Å. The first residue belongs to the protein and the second one to the peptide. SP2259 is the phosphorylated (doubly unprotonated) SER259. The tables were taken from the supporting information of ref. [32].

RSRSTpSTP (26 pairs)

Residue pair	Occupancy
ARG132-NH2 – SP2259-OE2	100.00
TYR133-OH – SP2259-OE2	100.00
ARG132-NH1 – SP2259-OE1	99.90
ARG57-NH2 – SP2259-OE1	99.88
ASN178-OD1 – THR260-N	99.26
ARG57-NH1 – SP2259-OE	98.98
LYS50-NZ – SP2259-OE	93.13
ASN229-ND2 – THR258-O	90.50
TRP233-NE1 – SER257-OG	76.98
ASN178-OD1 – THR260-OG1	57.33
GLU185-OE2 – SER257-N	53.07
GLU185-OE2 – SER257-OG	51.35
ARG57-NH1 – SP2259-OE1	44.01
ASN229-OD1 – THR258-N	40.30
GLU185-OE1 – SER257-N	38.87
GLU185-OE1 – SER257-OG	38.29
LYS125-NZ – THR260-OG1	35.52
GLU185-OE2 – ARG254-NH2	32.64
GLU185-OE1 – ARG254-NH2	30.01
GLU185-OE1 – ARG254-NH1	28.30
ASN178-ND2 – THR260-OG1	28.18
GLU185-OE2 – ARG254-NH1	22.24
ARG132-NH1 – SP2259-OE2	20.77
LYS50-NZ – SP2259-OE2	13.02

RSRSTSTP (20 pairs)

Residue pair	Occupancy
ASN178-OD1 – THR260-N	88.67
ASN229-ND2 – THR258-O	82.60
ASN178-ND2 – THR260-O	66.81
ASN229-OD1 – THR258-N	64.96
TRP233-NE1 – SER257-OG	55.87
LYS125-NZ – THR260-O	39.98
ARG132-NH1 – SER259-OG	39.74
SER46-OG – PRO261-O	26.35
SER46-OG – PRO261-OXT	25.91
LYS125-NZ – PRO261-O	25.13
ASN178-OD1 – THR260-OG1	25.03
GLU185-OE2 – SER257-N	24.98
LYS125-NZ – THR260-OG1	22.14
LYS125-NZ – PRO261-OXT	22.00
GLU185-OE1 – SER257-N	21.79
GLU185-OE2 – SER257-OG	20.71
ARG132-NH2 – SER259-OG	19.64
LYS50-NZ – PRO261-O	19.47
LYS50-NZ – PRO261-OXT	18.84
ASN178-ND2 – THR260-OG1	18.21
GLU185-OE1 – SER257-OG	17.30
GLU185-OE1 – SER255-OG	13.82
GLU185-OE2 – ARG256-NH2	12.34
GLU185-OE2 – ARG256-N	11.97
GLU185-OE1 – ARG256-NE	10.83
GLU136-OE1 – ARG256-NH2	10.67
GLU136-OE2 – ARG256-NH2	10.64
GLU136-OE2 – ARG256-NH1	10.16

TABLE A.11: Hydrogen-bond occupancy for hydrogen bonds between a 14-3-3η dimer and the RSRSTSTP or RSRSTpSTP peptide which exist in more than 10 % of all simulation frames. The first residue belongs to the protein and the second to the peptide. The tables were taken from the supporting information of ref. [32].

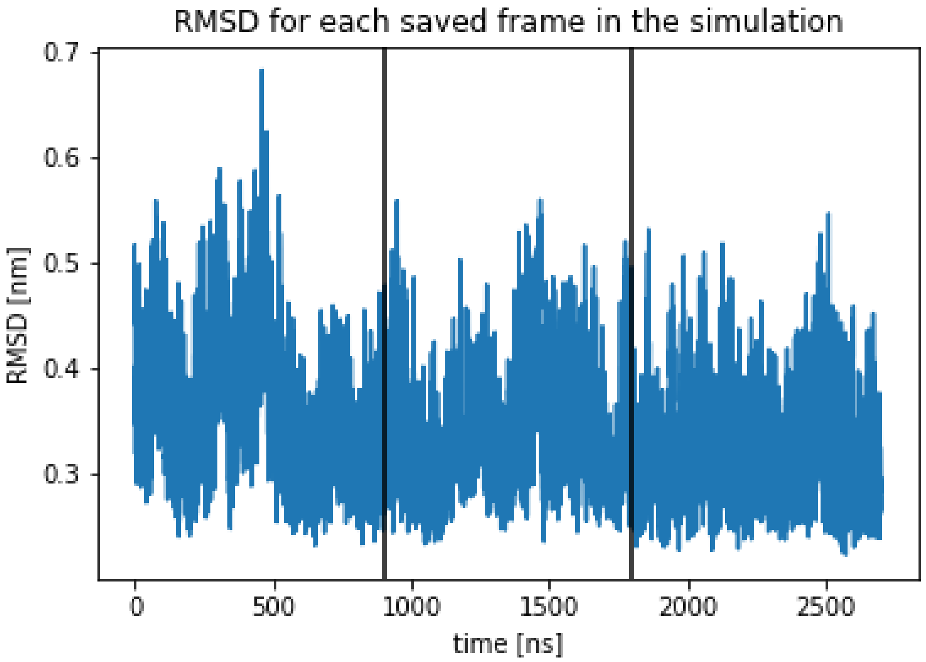


FIGURE A.7: Root mean squared deviations from the crystal structure during the courses of three combined simulations of a 14-3-3η dimer with a single RSRSTpSTP peptide bound to one monomer. The binding is stable during the course of all three simulations. The figure was taken from the supporting information of ref. [32].

	free energy difference	RSRSTpSTP (SP2)/ RSRSTSTP	RSRSTpSTP (SP1)/ RSRSTSTP	RSRSTpSTPNV (SP2)/ RSRSTSTPNV	RLYHpSLP (SP2)/ RLYHSLP
	$\Delta\Delta G^{\text{bound,SER}\to\text{SP1/SP2}}$	-1193.99 ± 0.76	-701.01 ± 3.80	-1188.99 ± 0.81	-1200.94 ± 2.67
−	$\Delta\Delta G^{\text{free,SER}\to\text{SP1/SP2}}$	-1167.08 ± 3.02	-695.66 ± 0.32	-1162.04 ± 2.86	-1163.65 ± 1.65
+ Corrections	$\Delta\Delta G^{\text{bound,SER}\to\text{SP1/SP2}}_{\text{electrostatictotal}}$	135.13	75.41	151.47	152.19
	$\Delta\Delta G^{\text{free,SER}\to\text{SP1/SP2}}_{\text{electrostatictotal}}$	149.54	79.99	159.66	162.34
	$\Delta\Delta G^{\text{bound,SER}\to\text{SP1/SP2}}_{\text{standardstate}}$	-16.03	-16.03	-16.03	-16.03
	$\Delta\Delta G^{\text{free,SER}\to\text{SP1/SP2}}_{\text{standardstate}}$	-11.12	-11.12	-12.06	-11.18
=	$\Delta\Delta G^{\text{bind,SP1/SP2}-\text{SER}}$	-41.32 ± 6.42	-14.84 ± 5.27	-39.11 ± 6.17	-50.80 ± 7.32

TABLE A.12: Binding free energy differences calculated from alchemical transformation simulations shown in table 3.5. All values are in kJ mol^{-1}. The binding free energy $\Delta\Delta G^{\text{bind}}$ is the difference of bound and free energy differences with the added corrections and is always given as the difference of the binding free energy of the phosphorylated minus the one of the unphosphorylated peptide. The table was taken from the supporting information of ref. [32].

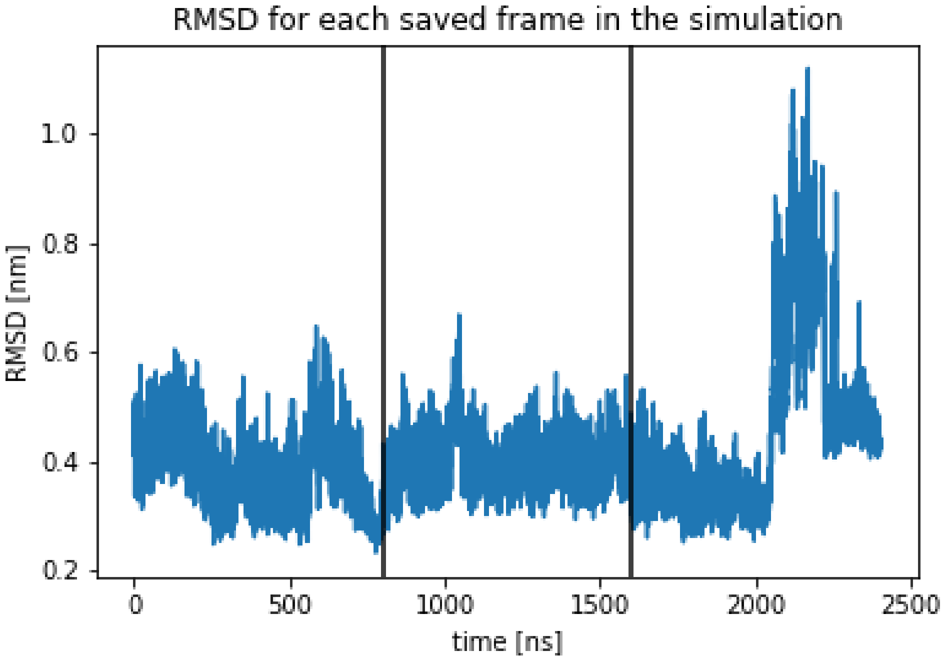


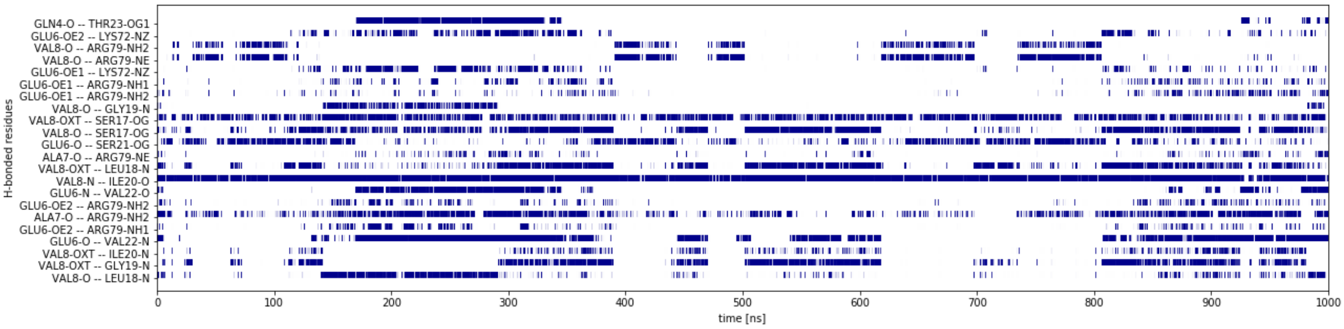
FIGURE A.8: Root mean squared deviations from the crystal structure during the courses of three combined simulations of a 14-3-3η dimer with a RSRSTpSTP and a HRYSTP peptide bound to one monomer each. During the course of the third simulation, the HRYSTP peptide is unbinding. The figure was taken from the supporting information of ref. [32].

Appendix B

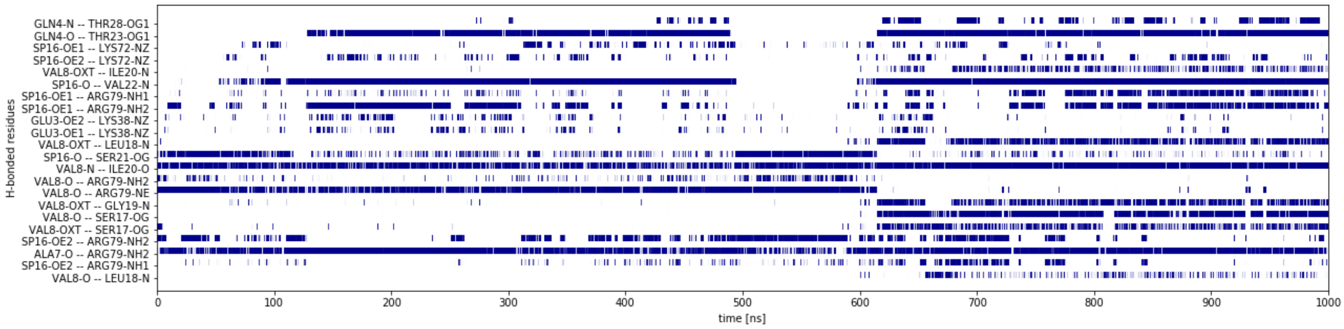
Supplementary Material for chapter 4

This supporting information belongs to chapter 4.

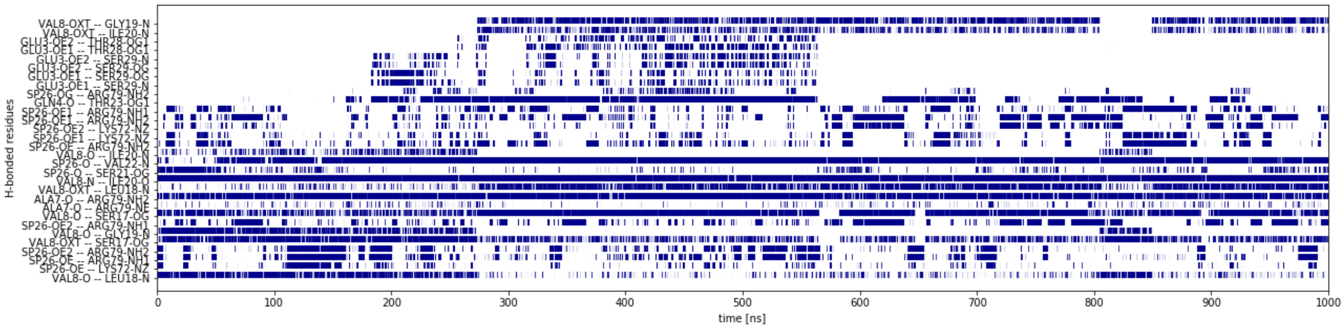
B.1 Hydrogen Bonds



(A) hPTP1E PDZ2 binding EQVEAV

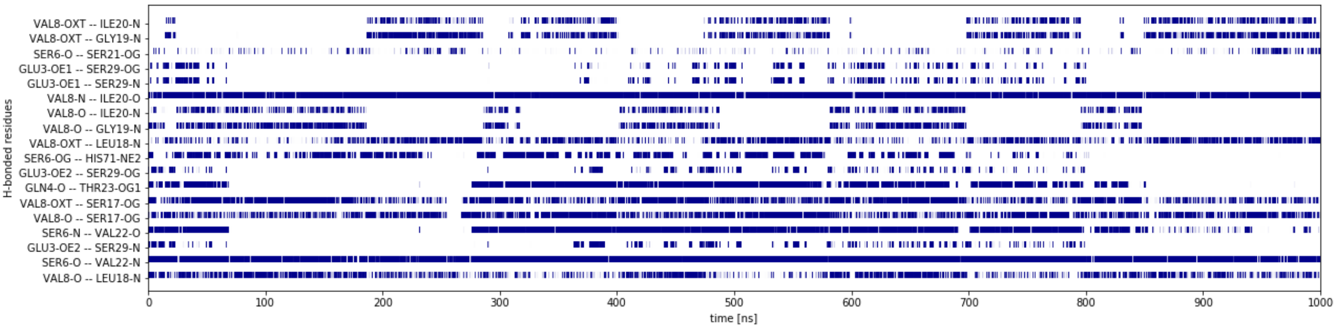


(B) hPTP1E PDZ2 binding EQVpSAV with singly charged phosphate

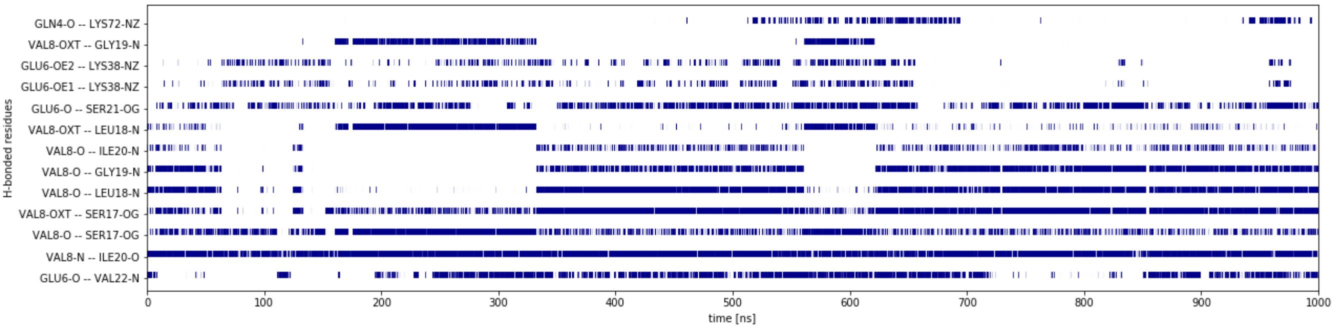


(C) hPTP1E PDZ2 binding EQVpSAV with doubly charged phosphate

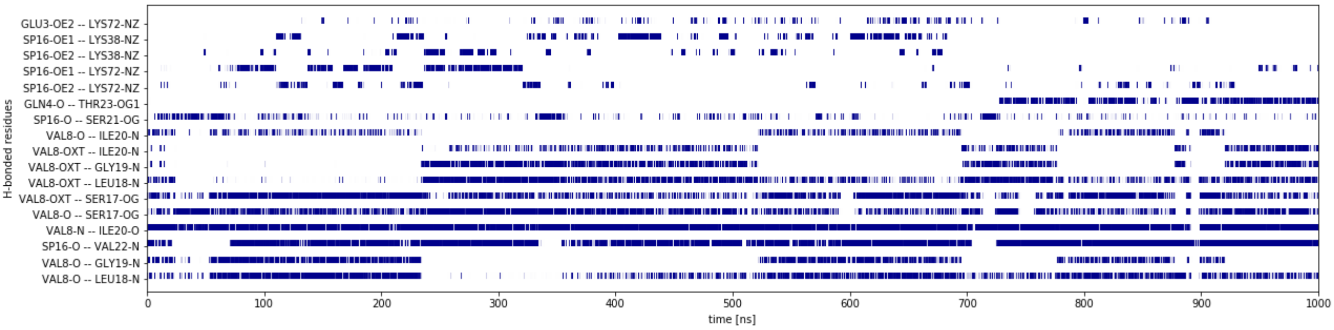
FIGURE B.1: Hydrogen bonds that exist in more than 10 % of all frames (white = no hydrogen bond, blue = hydrogen bond) between the PDZ2 domain of hPTP1E and (a) the EQVEAV peptide, (b) the phosphorylated EQVpSAV peptide with singly charged phosphate (SP1) and (c) the phosphorylated EQVpSAV peptide with doubly charged phosphate (SP2).



(A) hPTP1E PDZ2 R79A binding EQVSAV

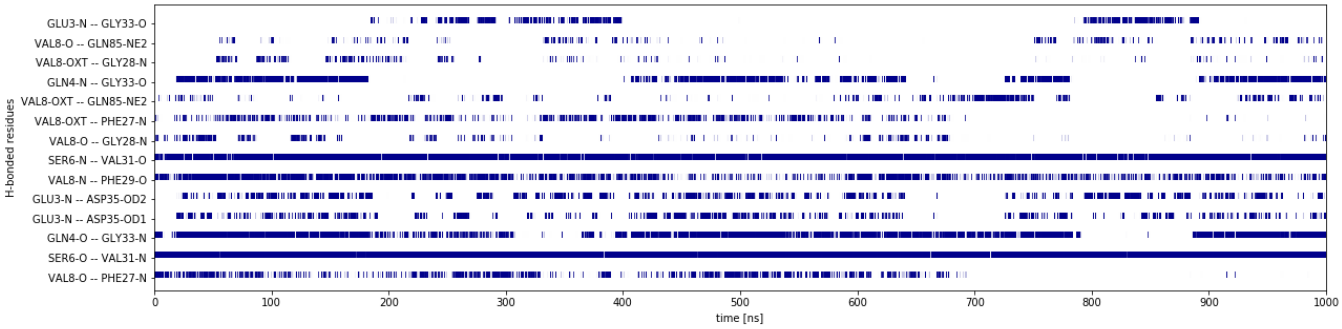


(B) hPTP1E PDZ2 R79A binding EQVEAV

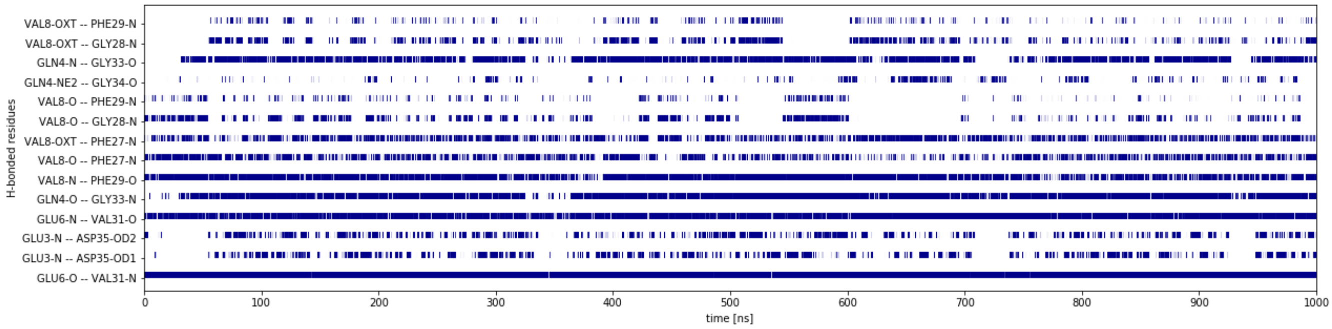


(C) hPTP1E PDZ2 R79A binding EQVpSAV with singly charged phosphate

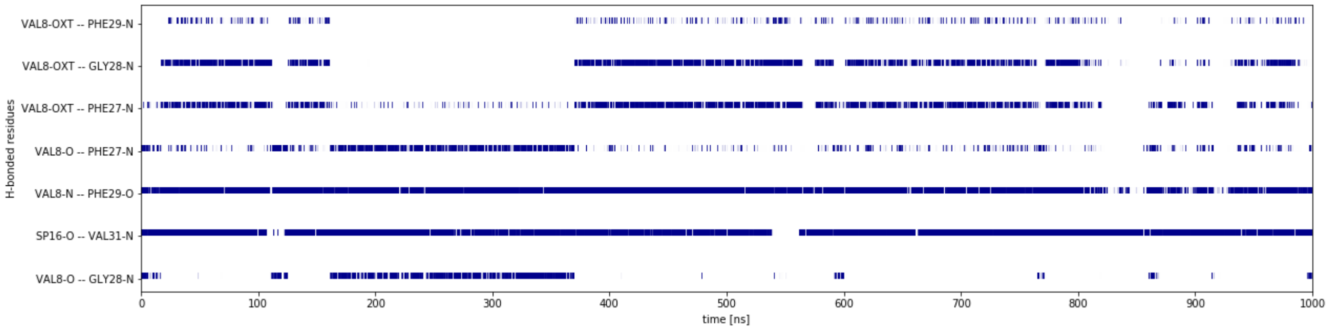
FIGURE B.2: Hydrogen bonds that exist in more than 10 % of all frames (white = no hydrogen bond, blue = hydrogen bond) between the PDZ2 R79A domain of hPTP1E and (a) the EQVSAV peptide, (b) the EQVEAV peptide and (c) the phosphorylated EQVpSAV peptide with singly charged phosphate (SP1).



(A) MAGI1 PDZ1 binding EQVSAV

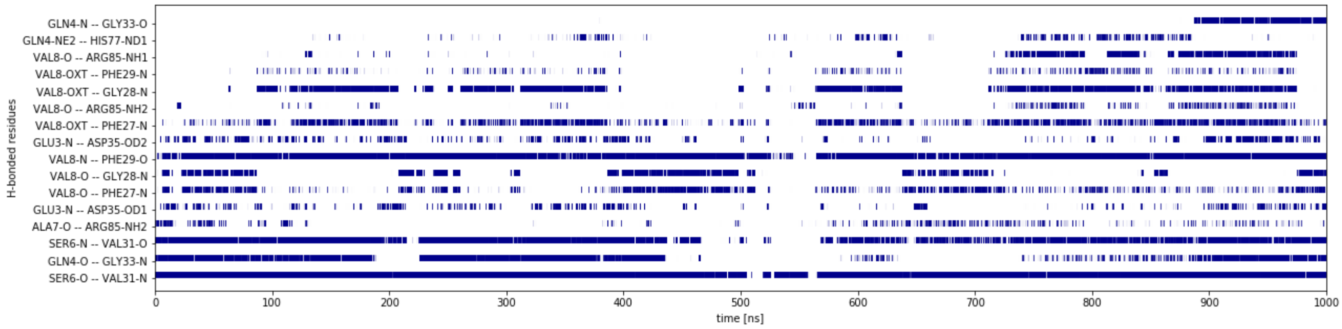


(B) MAGI1 PDZ1 binding EQVEAV

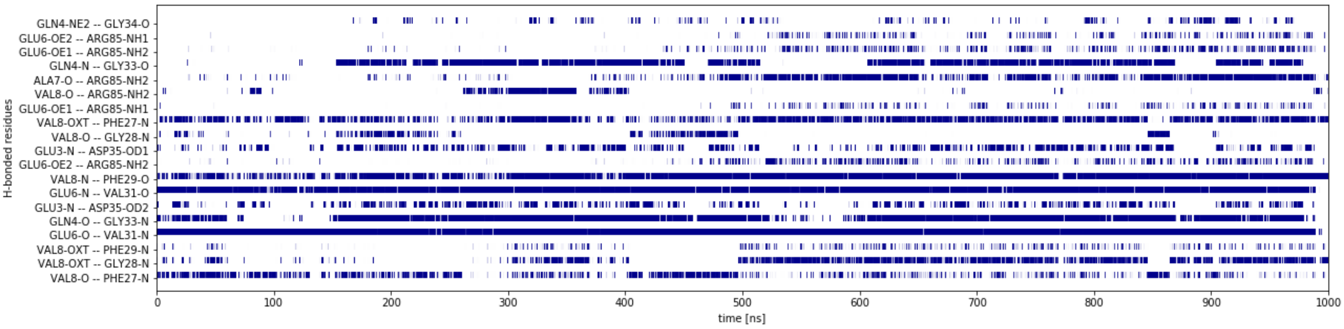


(C) MAGI1 PDZ1 binding EQVpSAV with singly charged phosphate

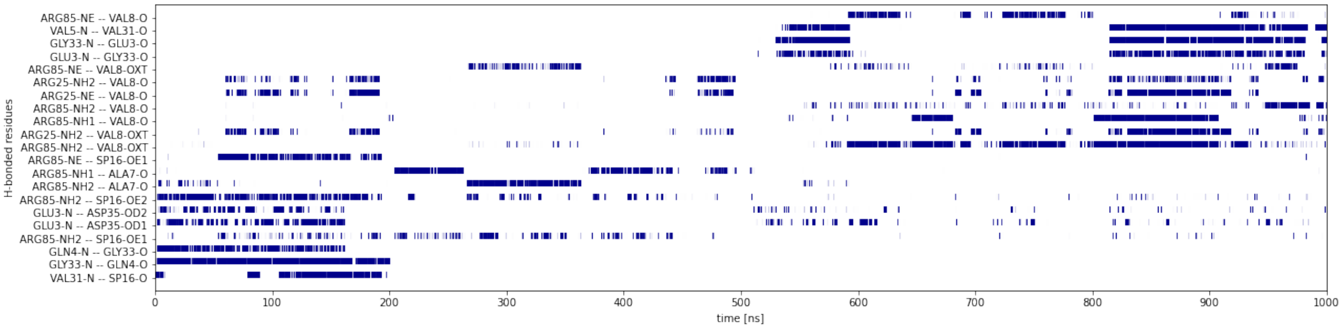
FIGURE B.3: Hydrogen bonds that exist in more than 10 % of all frames (white = no hydrogen bond, blue = hydrogen bond) between the PDZ1 domain of MAGI1 and (a) the EQVSAV peptide, (b) the EQVEAV peptide and (c) the phosphorylated EQVpSAV peptide with singly charged phosphate (SP1).



(A) MAGI1 PDZ1 Q85R binding EQVSAV



(B) MAGI1 PDZ1 Q85R binding EQVEAV



(C) MAGI1 PDZ1 Q85R binding EQVpSAV with singly charged phosphate

FIGURE B.4: Hydrogen bonds that exist in more than 10 % of all frames (white = no hydrogen bond, blue = hydrogen bond) between the PDZ1 Q85R domain of MAGI1 and (a) the EQVSAV peptide, (b) the EQVEAV peptide and (c) the phosphorylated EQVpSAV peptide with singly charged phosphate (SP1).

hPTP1E PDZ2 with EQVEAV

Interacting atoms	Occupancy [%]	Distance (STD) [Å]
VAL8-N – ILE20-O	96.21	2.98 (0.23)
VAL8-OXT – SER17-OG	67.73	3.17 (0.74)
VAL8-O – SER17-OG	51.93	3.69 (1.18)
ALA7-O – ARG79-NH2	49.85	3.84 (1.53)
GLU6-O – VAL22-N	49.18	3.95 (1.13)
VAL8-OXT – LEU18-N	48.17	3.59 (0.79)
GLU6-O – SER21-OG	40.43	3.55 (0.78)
VAL8-OXT – GLY19-N	39.43	4.04 (1.13)
VAL8-O – ARG79-NE	27.50	5.14 (1.71)
VAL8-O – LEU18-N	23.63	4.48 (1.37)
GLU6-N – VAL22-O	23.13	5.75 (2.13)
VAL8-O – ARG79-NH2	19.32	5.61 (1.87)
GLN4-O – THR23-OG1	18.56	8.81 (3.83)
GLU6-OE2 – LYS72-NZ	18.53	8.87 (4.86)
VAL8-OXT – ILE20-N	18.14	4.65 (1.10)
GLU6-OE1 – LYS72-NZ	17.63	8.86 (4.84)
GLU6-OE2 – ARG79-NH1	13.49	7.24 (3.47)
GLU6-OE2 – ARG79-NH2	13.05	7.00 (3.44)
VAL8-O – GLY19-N	13.04	5.26 (1.55)
GLU6-OE1 – ARG79-NH1	12.27	7.23 (3.46)
GLU6-OE1 – ARG79-NH2	11.58	7.00 (3.42)
ALA7-O – ARG79-NE	10.40	4.53 (1.09)

TABLE B.1: Percentage of hydrogen-bond occupancies and the distances between the involved heavy atoms for hydrogen bonds between hPTP1E PDZ2 and the EQVEAV peptide, existing in 10 % or more of all simulation frames. In the left column, first the peptide residue and second the protein residue is given.

hPTP1E PDZ2 with EQVpSAV
(singly charged (SP1) phosphate)

Interacting atoms	**Occupancy [%]**	**Distance (STD) [Å]**
VAL8-N – ILE20-O	88.78	3.10 (0.24)
ALA7-O – ARG79-NH2	88.46	2.84 (0.26)
SP16-O – VAL22-N	80.85	3.18 (0.74)
GLN4-O – THR23-OG1	72.07	4.43 (3.08)
VAL8-O – ARG79-NE	57.85	4.00 (1.41)
SP16-OE1 – ARG79-NH2	48.91	3.92 (1.25)
SP16-OE2 – ARG79-NH2	39.22	4.12 (1.18)
SP16-O – SER21-OG	39.05	3.50 (0.71)
VAL8-O – SER17-OG	33.99	5.17 (1.94)
VAL8-OXT – LEU18-N	30.48	4.45 (1.16)
VAL8-OXT – GLY19-N	27.83	4.40 (1.10)
SP16-OE1 – ARG79-NH1	26.87	4.38 (1.35)
VAL8-OXT – SER17-OG	25.55	4.11 (1.15)
GLN4-N – THR28-OG1	18.70	7.55 (4.55)
VAL8-OXT – ILE20-N	18.47	4.88 (1.11)
SP16-OE2 – LYS72-NZ	18.19	6.09 (2.59)
SP16-OE2 – ARG79-NH1	14.49	4.48 (1.14)
SP16-OE1 – LYS72-NZ	13.78	6.29 (2.74)
VAL8-O – ARG79-NH2	13.76	4.54 (1.31)
GLU3-OE2 – LYS38-NZ	11.84	8.36 (3.75)
VAL8-O – LEU18-N	10.95	5.76 (1.73)
GLU3-OE1 – LYS38-NZ	10.75	8.38 (3.72)

TABLE B.2: Percentage of hydrogen-bond occupancies and the distances between the involved heavy atoms for hydrogen bonds between hPTP1E PDZ2 and EQVpSAV peptide with singly charged (SP1) phosphate, existing in 10 % or more of all simulation frames. In the left column, first the peptide residue and second the protein residue is given.

hPTP1E PDZ2 with EQVpSAV
(doubly charged (SP2) phosphate)

Interacting atoms	Occupancy [%]	Distance (STD) [Å]
VAL8-N – ILE20-O	98.51	2.94 (0.18)
ALA7-O – ARG79-NH2	93.23	2.87 (0.24)
SP26-O – VAL22-N	93.00	2.96 (0.32)
VAL8-O – SER17-OG	83.71	2.94 (0.50)
VAL8-OXT – SER17-OG	69.93	3.04 (0.46)
VAL8-OXT – LEU18-N	64.47	3.25 (0.44)
GLN4-O – THR23-OG1	55.96	4.67 (2.43)
VAL8-OXT – GLY19-N	52.93	3.51 (0.83)
VAL8-O – LEU18-N	42.55	3.47 (0.48)
SP26-OE2 – ARG79-NH1	39.93	3.72 (0.96)
SP26-OE1 – ARG79-NH2	39.23	3.72 (0.94)
VAL8-OXT – ILE20-N	36.07	4.15 (1.06)
SP26-OE – ARG79-NH1	34.96	3.84 (0.96)
SP26-OE – ARG79-NH2	33.20	3.85 (0.93)
SP26-OE1 – ARG79-NH1	32.79	3.78 (0.92)
SP26-OE2 – ARG79-NH2	32.49	3.76 (0.88)
VAL8-O – GLY19-N	26.12	4.07 (0.86)
SP26-OE2 – LYS72-NZ	24.59	4.90 (2.22)
SP26-OE1 – LYS72-NZ	21.19	5.01 (2.22)
SP26-O – SER21-OG	19.88	3.88 (0.72)
SP26-OE – LYS72-NZ	16.95	4.96 (2.11)
VAL8-O – ILE20-N	14.81	4.90 (1.03)
GLU3-OE1 – THR28-OG1	14.04	10.34 (5.91)
GLU3-OE2 – SER29-N	13.95	10.84 (6.39)
GLU3-OE1 – SER29-N	13.27	10.85 (6.38)
ALA7-O – ARG79-NE	12.32	3.65 (0.37)
GLU3-OE2 – SER29-OG	12.06	11.34 (6.66)
GLU3-OE2 – THR28-OG1	11.96	10.37 (5.87)
GLU3-OE1 – SER29-OG	11.33	11.38 (6.61)
SP26-OG – ARG79-NH2	11.01	3.91 (0.41)

TABLE B.3: Percentage of hydrogen-bond occupancies and the distances between the involved heavy atoms for hydrogen bonds between hPTP1E PDZ2 and EQVpSAV peptide with doubly charged (SP2) phosphate, existing in 10 % or more of all simulation frames. In the left column, first the peptide residue and second the protein residue is given.

hPTP1E PDZ2 R79A with EQVSAV

Interacting atoms	Occupancy [%]	Distance (STD) [Å]
SER6-O – VAL22-N	97.63	2.91 (0.26)
VAL8-N – ILE20-O	96.75	2.97 (0.21)
VAL8-O – SER17-OG	72.61	3.09 (0.69)
VAL8-OXT – SER17-OG	71.61	3.05 (0.59)
SER6-N – VAL22-O	63.38	3.63 (1.06)
VAL8-O – LEU18-N	57.36	3.40 (0.67)
GLN4-O – THR23-OG1	55.33	5.45 (3.73)
VAL8-OXT – LEU18-N	55.13	3.39 (0.59)
SER6-OG – HIS71-NE2	42.13	5.07 (3.11)
VAL8-OXT – GLY19-N	40.45	3.78 (0.96)
VAL8-O – GLY19-N	33.81	3.95 (1.02)
VAL8-OXT – ILE20-N	32.00	4.43 (1.16)
VAL8-O – ILE20-N	24.91	4.66 (1.18)
GLU3-OE1 – SER29-N	16.83	10.55 (7.37)
GLU3-OE2 – SER29-N	16.59	10.57 (7.38)
GLU3-OE1 – SER29-OG	14.75	11.31 (8.16)
GLU3-OE2 – SER29-OG	13.86	11.34 (8.16)
SER6-O – SER21-OG	12.18	3.85 (0.63)

TABLE B.4: Percentage of hydrogen-bond occupancies and the distances between the involved heavy atoms for hydrogen bonds between hPTP1E PDZ2 R79A and the EQVSAV peptide, existing in 10 % or more of all simulation frames. In the left column, first the peptide residue and second the protein residue is given.

hPTP1E PDZ2 R79A with EQVEAV

Interacting atoms	Occupancy [%]	Distance (STD) [Å]
VAL8-N – ILE20-O	94.50	3.02 (0.23)
VAL8-OXT – SER17-OG	78.27	3.04 (0.70)
VAL8-O – SER17-OG	66.21	3.04 (0.44)
VAL8-O – LEU18-N	64.94	3.31 (0.67)
VAL8-O – GLY19-N	58.11	3.53 (0.98)
GLU6-O – SER21-OG	57.57	3.37 (0.83)
GLU6-O – VAL22-N	55.94	3.77 (1.03)
VAL8-OXT – LEU18-N	30.68	3.79 (0.92)
VAL8-O – ILE20-N	30.08	4.20 (1.07)
GLU6-OE1 – LYS38-NZ	23.52	6.38 (3.54)
GLU6-OE2 – LYS38-NZ	22.17	6.40 (3.54)
VAL8-OXT – GLY19-N	19.84	4.42 (1.14)
GLN4-O – LYS72-NZ	14.32	7.39 (4.10)

TABLE B.5: Percentage of hydrogen-bond occupancies and the distances between the involved heavy atoms for hydrogen bonds between hPTP1E PDZ2 R79A and the EQVEAV peptide, existing in 10 % or more of all simulation frames. In the left column, first the peptide residue and second the protein residue is given.

hPTP1E PDZ2 R79A with EQVpSAV (singly charged (SP1) phosphate)

Interacting atoms	Occupancy [%]	Distance (STD) [Å]
VAL8-N – ILE20-O	96.40	2.98 (0.24)
SP16-O – VAL22-N	84.85	3.18 (0.66)
VAL8-O – SER17-OG	70.81	3.10 (0.65)
VAL8-OXT – SER17-OG	70.36	3.12 (0.66)
VAL8-O – LEU18-N	61.93	3.31 (0.52)
VAL8-OXT – LEU18-N	55.58	3.40 (0.64)
VAL8-O – GLY19-N	40.87	3.79 (0.95)
VAL8-OXT – GLY19-N	38.21	3.89 (1.02)
VAL8-O – ILE20-N	26.32	4.52 (1.15)
VAL8-OXT – ILE20-N	23.54	4.64 (1.18)
SP16-O – SER21-OG	21.86	3.66 (0.67)
GLN4-O – THR23-OG1	19.48	8.49 (3.56)
SP16-OE1 – LYS38-NZ	18.73	8.74 (4.47)
SP16-OE1 – LYS72-NZ	17.66	7.84 (4.12)
SP16-OE2 – LYS72-NZ	12.51	7.62 (3.87)
SP16-OE2 – LYS38-NZ	11.17	8.80 (4.26)
GLU3-OE2 – LYS72-NZ	10.95	9.46 (4.74)

TABLE B.6: Percentage of hydrogen-bond occupancies and the distances between the involved heavy atoms for hydrogen bonds between hPTP1E PDZ2 R79A and EQVpSAV peptide with singly charged (SP1) phosphate, existing in 10 % or more of all simulation frames. In the left column, first the peptide residue and second the protein residue is given.

MAGI1 PDZ1 with EQVSAV

Interacting atoms	Occupancy [%]	Distance (STD) [Å]
SER6-O – VAL31-N	99.74	2.88 (0.13)
SER6-N – VAL31-O	98.12	2.97 (0.18)
GLN4-O – GLY33-N	73.13	3.44 (0.91)
VAL8-N – PHE29-O	62.45	3.31 (0.41)
GLN4-N – GLY33-O	46.04	4.19 (1.43)
GLU3-N – ASP35-OD2	39.61	4.92 (2.85)
VAL8-O – PHE27-N	35.02	4.18 (1.25)
GLU3-N – ASP35-OD1	34.38	5.02 (2.82)
VAL8-OXT – PHE27-N	29.26	4.30 (1.24)
VAL8-OXT – GLN85-NE2	18.55	5.58 (1.85)
GLU3-N – GLY33-O	18.34	5.01 (1.93)
VAL8-O – GLY28-N	13.40	5.06 (1.38)
VAL8-O – GLN85-NE2	11.84	5.79 (1.75)
VAL8-OXT – GLY28-N	10.18	5.28 (1.37)

TABLE B.7: Percentage of hydrogen-bond occupancies and the distances between the involved heavy atoms for hydrogen bonds between MAGI1 PDZ1 and EQVSAV peptide, existing in 10 % or more of all simulation frames. In the left column, first the peptide residue and second the protein residue is given.

MAGI1 PDZ1 with EQVEAV

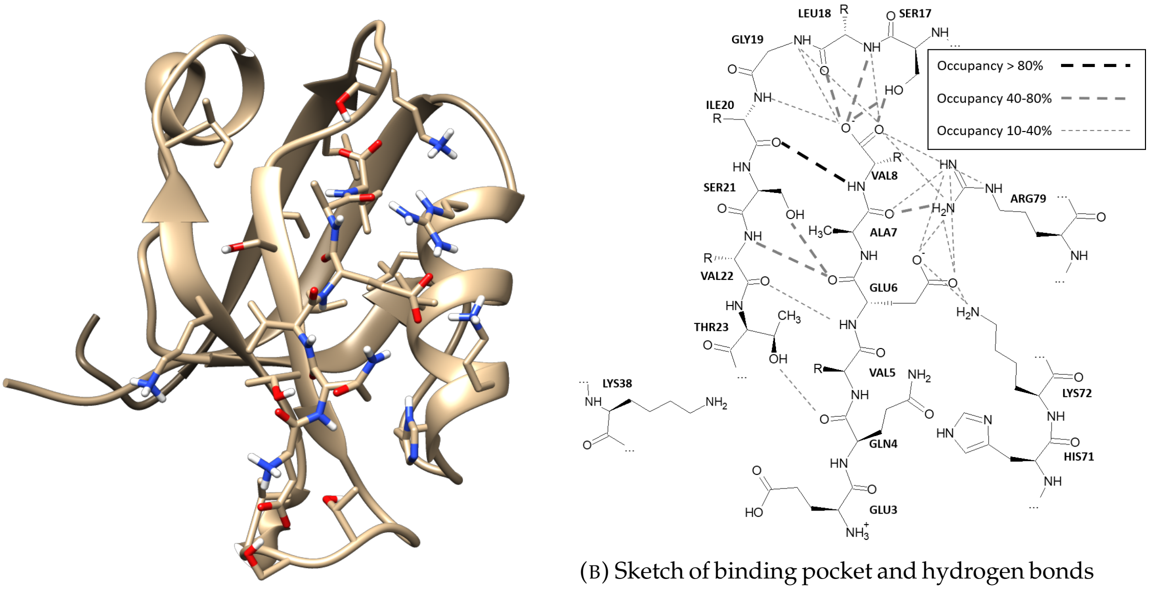
Interacting atoms	Occupancy [%]	Distance (STD) [Å]
GLU6-O – VAL31-N	99.79	2.87 (0.13)
GLU6-N – VAL31-O	95.03	3.07 (0.18)
GLN4-O – GLY33-N	89.79	3.09 (0.60)
VAL8-N – PHE29-O	87.51	3.08 (0.29)
GLN4-N – GLY33-O	77.66	3.38 (1.08)
VAL8-OXT – PHE27-N	64.41	3.23 (0.54)
VAL8-O – PHE27-N	60.50	3.27 (0.57)
GLU3-N – ASP35-OD1	41.65	4.48 (2.39)
GLU3-N – ASP35-OD2	41.27	4.51 (2.41)
VAL8-OXT – GLY28-N	36.29	3.94 (1.04)
VAL8-O – GLY28-N	33.58	4.05 (1.08)
GLN4-NE2 – GLY34-O	15.23	5.56 (1.87)
VAL8-OXT – PHE29-N	14.89	4.80 (1.21)
VAL8-O – PHE29-N	14.75	4.93 (1.24)

TABLE B.8: Percentage of hydrogen-bond occupancies and the distances between the involved heavy atoms for hydrogen bonds between MAGI1 PDZ1 and EQVEAV peptide, existing in 10 % or more of all simulation frames. In the left column, first the peptide residue and second the protein residue is given.

MAGI1 PDZ1 with EQVpSAV
(singly charged (SP1) phosphate)

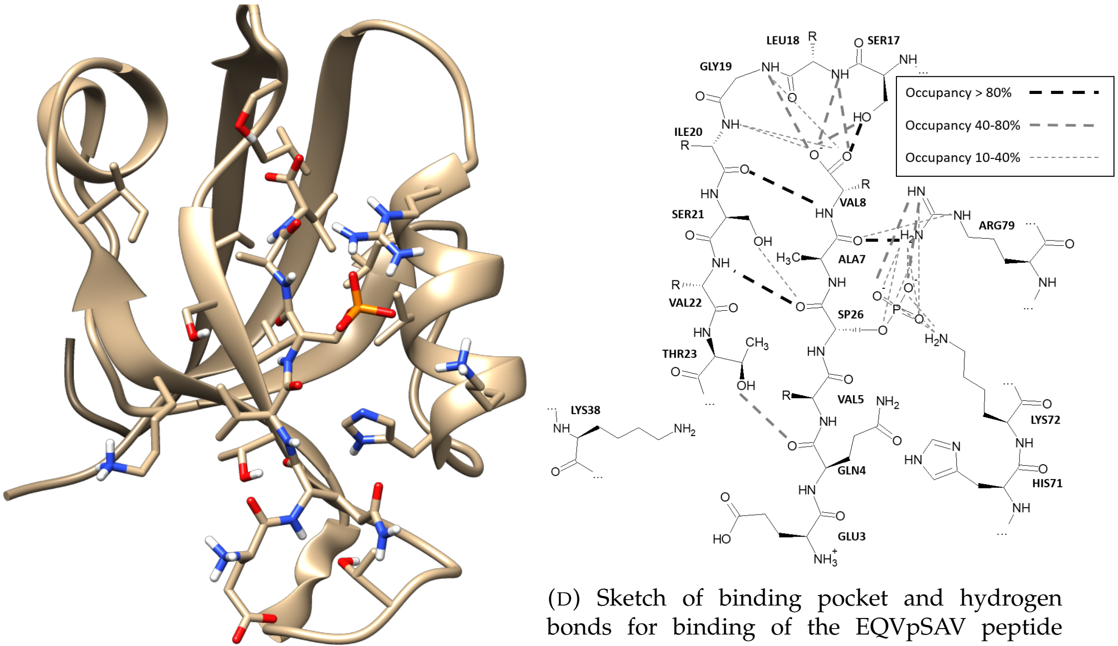
Interacting atoms	Occupancy [%]	Distance (STD) [Å]
SP16-O – VAL31-N	94.60	3.03 (0.53)
VAL8-N – PHE29-O	92.92	2.99 (0.31)
VAL8-OXT – PHE27-N	54.44	3.48 (0.79)
VAL8-OXT – GLY28-N	52.27	3.73 (1.14)
VAL8-O – PHE27-N	34.86	3.75 (0.86)
VAL8-O – GLY28-N	22.86	4.46 (1.15)
VAL8-OXT – PHE29-N	16.79	4.49 (1.20)

TABLE B.9: Percentage of hydrogen-bond occupancies and the distances between the involved heavy atoms for hydrogen bonds between MAGI1 PDZ1 and EQVpSAV peptide with singly charged (SP1) phosphate, existing in 10 % or more of all simulation frames. In the left column, first the peptide residue and second the protein residue is given.



(A) hPTP1E PDZ2 binding EQVEAV

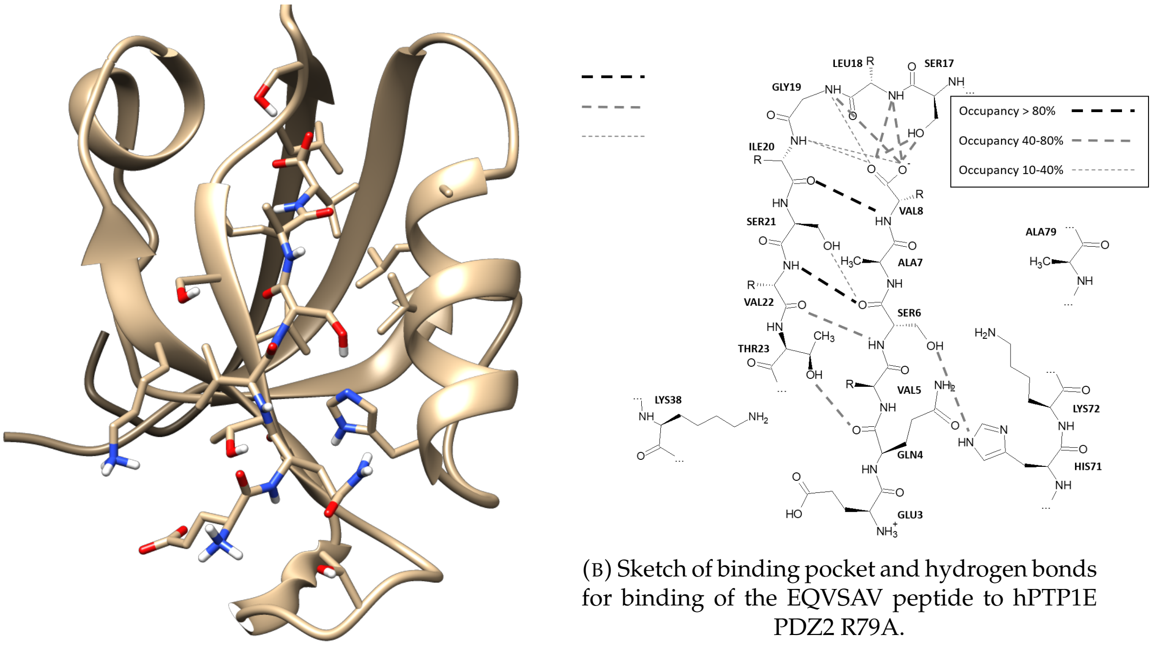
(B) Sketch of binding pocket and hydrogen bonds for binding of the EQVEAV peptide to hPTP1E PDZ2.



(C) hPTP1E PDZ2 binding EQVpSAV with doubly charged (SP2) phosphate

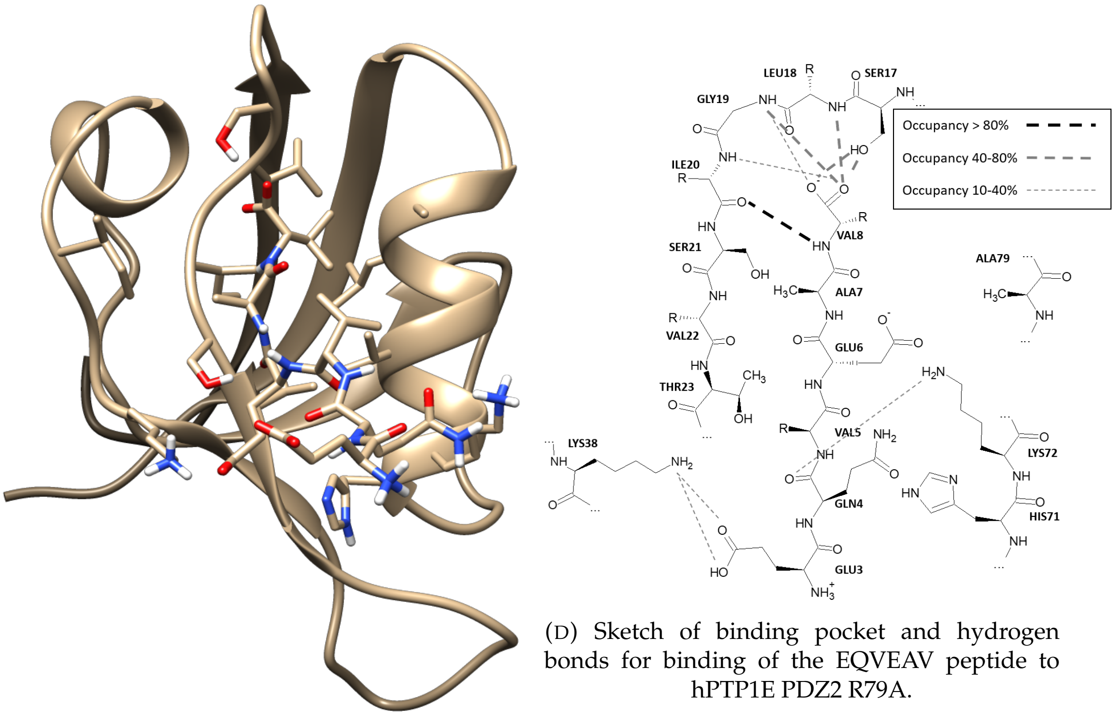
(D) Sketch of binding pocket and hydrogen bonds for binding of the EQVpSAV peptide with doubly charged (SP2) phosphate to hPTP1E PDZ2.

FIGURE B.5: Representative structures of the EQVEAV peptide (top left) and the EQVpSAV peptide with doubly charged (SP2) phosphate group (bottom left) binding the hPTP1E PDZ2 domain. These structures were extracted similar to our previous study [32] (see chapter 3) by finding the centroid, the frame with the highest sum of similarities, of all trajectory frames of 1 µs plain MD simulations with 2D sketches of the binding pocket and the relevant hydrogen bonds between protein and peptide on the right side, respectively. As a distance metric we used the pairwise RMSD and then used the pairwise distances to calculate a pairwise similarity.



(A) hPTP1E PDZ2 R79A binding EQVSAV

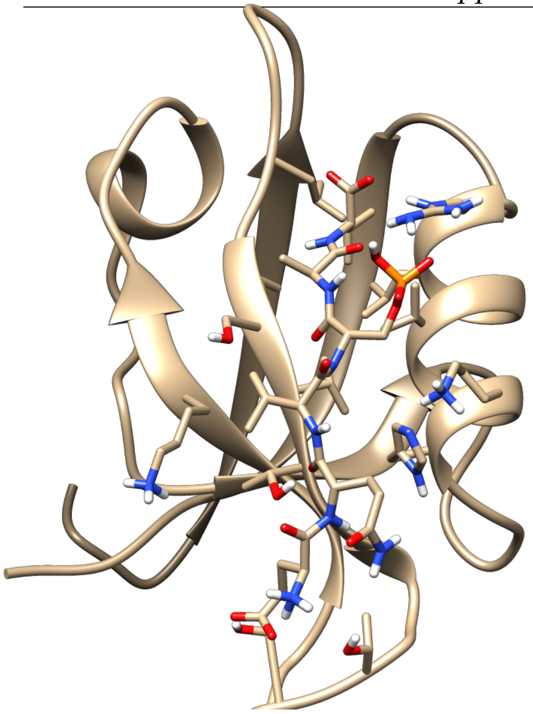
(B) Sketch of binding pocket and hydrogen bonds for binding of the EQVSAV peptide to hPTP1E PDZ2 R79A.



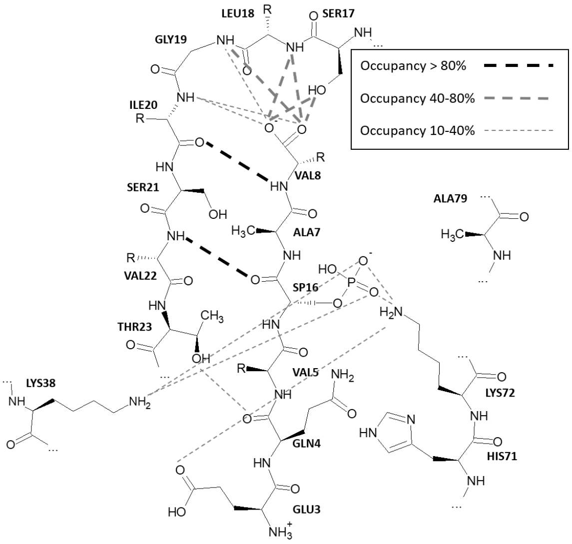
(D) Sketch of binding pocket and hydrogen bonds for binding of the EQVEAV peptide to hPTP1E PDZ2 R79A.

(C) hPTP1E PDZ2 R79A binding EQVEAV

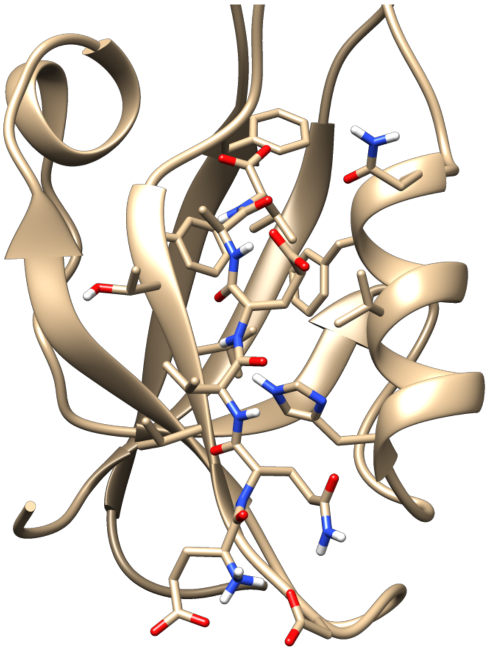
FIGURE B.6: Representative structures of the EQVSAV peptide (top left) and the EQVEAV peptide (bottom left) binding the hPTP1E PDZ2 R79A domain. These structures were extracted similar to our previous study [32] (see chapter 3) by finding the centroid, the frame with the highest sum of similarities, of all trajectory frames of 1 µs plain MD simulations with 2D sketches of the binding pocket and the relevant hydrogen bonds between protein and peptide on the right side, respectively. As a distance metric we used the pairwise RMSD and then used the pairwise distances to calculate a pairwise similarity.



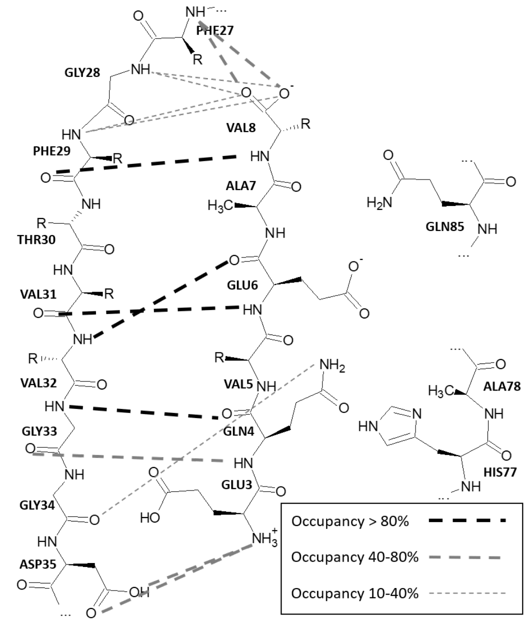
(A) hPTP1E PDZ2 R79A binding EQVpSAV with singly charged (SP1) phosphate group



(B) Sketch of binding pocket and hydrogen bonds for binding of the EQVpSAV peptide with singly charged (SP1) phosphate group to hPTP1E PDZ2 R79A.

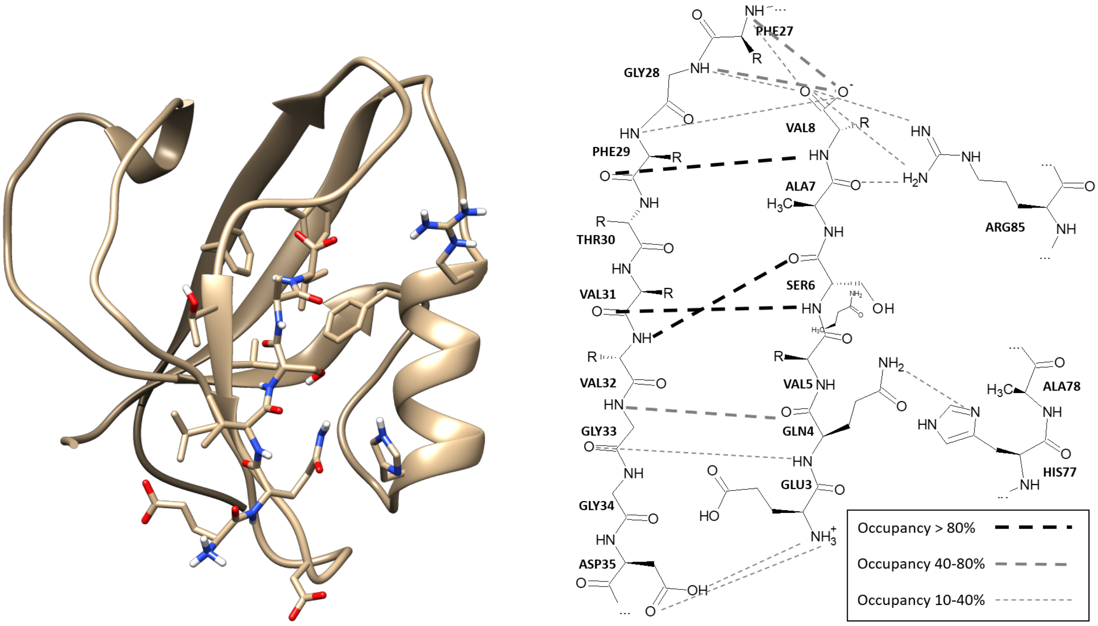


(C) MAGI1 PDZ1 binding EQVEAV



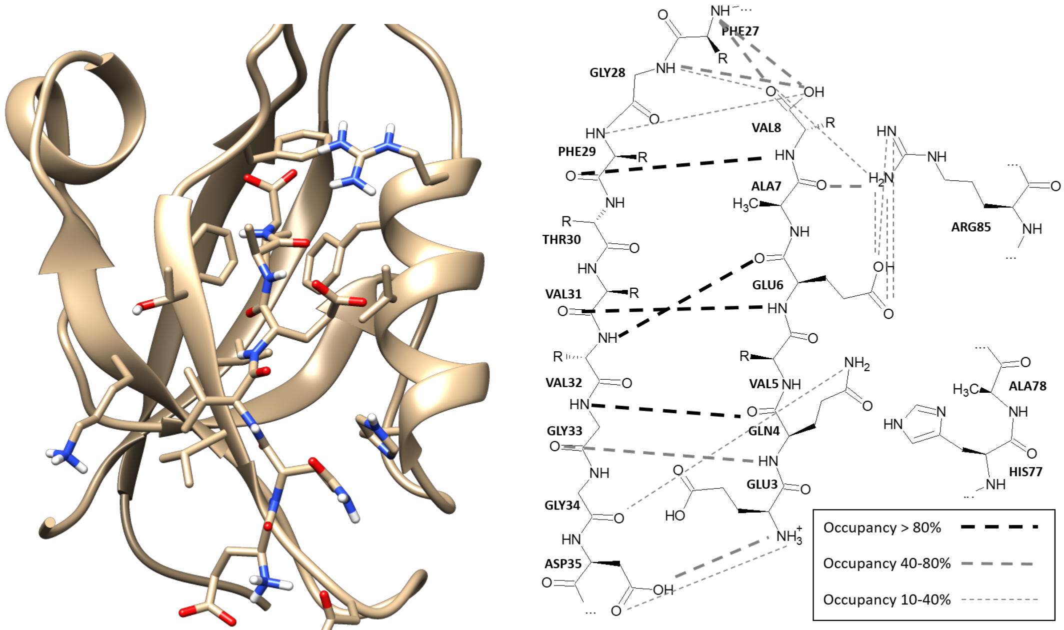
(D) Sketch of binding pocket and hydrogen bonds for binding of the EQVEAV to MAGI1 PDZ1.

FIGURE B.7: Representative structures of the EQVpSAV peptide with singly charged (SP1) phosphate group binding the hPTP1E PDZ2 R79A domain (top left) and the EQVEAV peptide binding the MAGI1 PDZ1 domain (bottom left). These structures were extracted similar to our previous study [32] (see chapter 3) by finding the centroid, the frame with the highest sum of similarities, of all trajectory frames of 1 µs plain MD simulations with 2D sketches of the binding pocket and the relevant hydrogen bonds between protein and peptide on the right side, respectively. As a distance metric we used the pairwise RMSD and then used the pairwise distances to calculate a pairwise similarity.



(A) MAGI1 PDZ1 Q85R binding EQVSAV

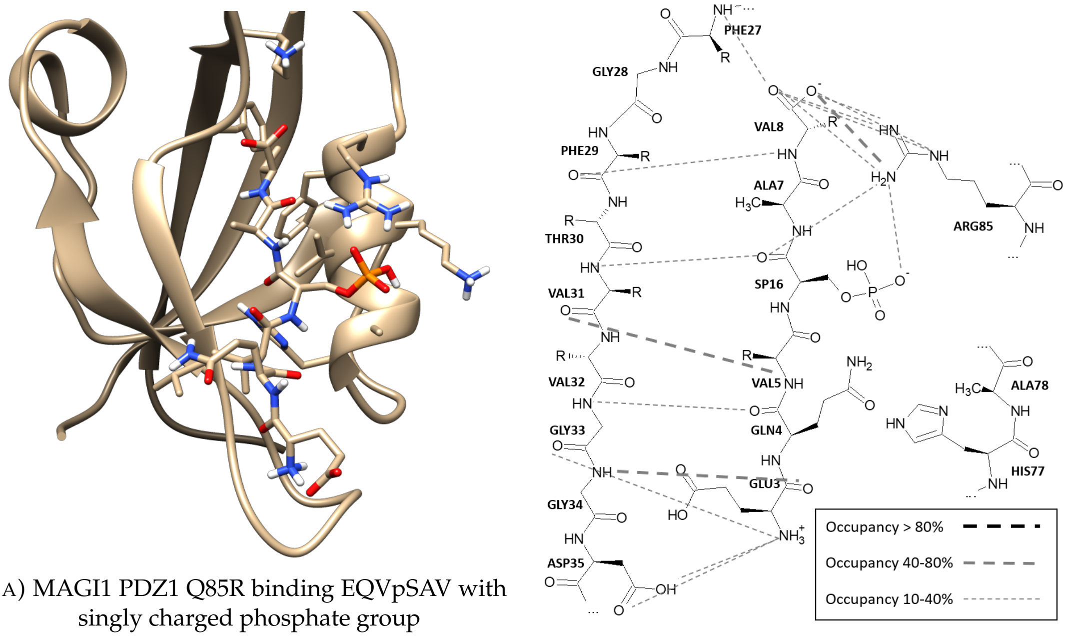
(B) Sketch of binding pocket and hydrogen bonds for binding of the EQVSAV to MAGI1 PDZ1 Q85R.



(C) MAGI1 PDZ1 Q85R binding EQVEAV

(D) Sketch of binding pocket and hydrogen bonds for binding of the EQVEAV to MAGI1 PDZ1 Q85R.

FIGURE B.8: Representative structures of the EQVSAV peptide (top left) and the EQVEAV peptide (bottom left) binding the MAGI1 PDZ1 Q85R domain. These structures were extracted similar to our previous study [32] (see chapter 3) by finding the centroid, the frame with the highest sum of similarities, of all trajectory frames of 1 µs plain MD simulations with 2D sketches of the binding pocket and the relevant hydrogen bonds between protein and peptide on the right side, respectively. As a distance metric we used the pairwise RMSD and then used the pairwise distances to calculate a pairwise similarity.



(A) MAGI1 PDZ1 Q85R binding EQVpSAV with singly charged phosphate group

(B) Sketch of binding pocket and hydrogen bonds for binding of the EQVpSAV with singly charged (SP1) phosphate group to MAGI1 PDZ1 Q85R.

FIGURE B.9: Representative structure of the EQVpSAV peptide with singly charged (SP1) phosphate group binding the MAGI1 PDZ1 Q85R domain (left). These structures were extracted similar to our previous study [32] (see chapter 3) by finding the centroid, the frame with the highest sum of similarities, of all trajectory frames of 1 µs plain MD simulations with 2D sketches of the binding pocket and the relevant hydrogen bonds between protein and peptide on the right side, respectively. As a distance metric we used the pairwise RMSD and then used the pairwise distances to calculate a pairwise similarity.

MAGI1 PDZ1 Q85R with EQVSAV

Interacting atoms	Occupancy [%]	Distance (STD) [Å]
SER6-O – VAL31-N	97.22	2.93 (0.44)
VAL8-N – PHE29-O	93.59	2.99 (0.42)
SER6-N – VAL31-O	81.92	3.25 (0.71)
GLN4-O – GLY33-N	56.40	4.27 (2.25)
VAL8-OXT – GLY28-N	55.23	3.79 (1.26)
VAL8-OXT – PHE27-N	50.74	3.54 (0.95)
VAL8-O – PHE27-N	39.85	3.72 (1.00)
VAL8-O – GLY28-N	30.73	4.26 (1.23)
GLU3-N – ASP35-OD2	25.64	6.26 (3.39)
GLU3-N – ASP35-OD1	21.86	6.40 (3.38)
VAL8-O – ARG85-NH1	20.45	6.48 (2.80)
ALA7-O – ARG85-NH2	16.19	5.76 (2.71)
VAL8-OXT – PHE29-N	15.82	4.65 (1.31)
GLN4-NE2 – HIS77-ND1	13.10	7.09 (3.21)
VAL8-O – ARG85-NH2	12.31	6.47 (2.48)
GLN4-N – GLY33-O	10.52	6.36 (2.66)

TABLE B.10: Percentage of hydrogen-bond occupancies and the distances between the involved heavy atoms for hydrogen bonds between MAGI1 PDZ1 Q85R and EQVSAV peptide, existing in 10 % or more of all simulation frames. In the left column, first the peptide residue and second the protein residue is given.

MAGI1 PDZ1 Q85R with EQVEAV

Interacting atoms	Occupancy [%]	Distance (STD) [Å]
GLU6-O – VAL31-N	98.95	2.89 (0.22)
GLU6-N – VAL31-O	96.18	3.04 (0.35)
VAL8-N – PHE29-O	87.55	3.04 (0.36)
GLN4-O – GLY33-N	80.42	3.35 (1.13)
VAL8-OXT – PHE27-N	66.47	3.34 (0.75)
GLN4-N – GLY33-O	61.15	4.11 (1.96)
VAL8-OXT – GLY28-N	46.28	3.92 (1.24)
VAL8-O – PHE27-N	43.99	3.64 (0.86)
ALA7-O – ARG85-NH2	42.91	4.29 (2.07)
GLU3-N – ASP35-OD1	40.91	4.88 (3.04)
GLU3-N – ASP35-OD2	37.64	4.96 (3.05)
GLU6-OE2 – ARG85-NH2	20.02	5.41 (2.41)
GLU6-OE1 – ARG85-NH2	19.72	5.43 (2.41)
VAL8-O – GLY28-N	14.99	4.69 (1.14)
VAL8-OXT – PHE29-N	14.67	4.67 (1.29)
GLN4-NE2 – GLY34-O	13.00	6.52 (2.80)
VAL8-O – ARG85-NH2	12.85	6.21 (2.11)
GLU6-OE1 – ARG85-NH1	11.29	5.85 (2.20)
GLU6-OE2 – ARG85-NH1	10.89	5.84 (2.19)

TABLE B.11: Percentage of hydrogen-bond occupancies and the distances between the involved heavy atoms for hydrogen bonds between MAGI1 PDZ1 Q85R and EQVEAV peptide, existing in 10 % or more of all simulation frames. In the left column, first the peptide residue and second the protein residue is given.

MAGI1 PDZ1 Q85R with EQVpSAV
(singly charged (SP1) phosphate)

Interacting atoms	Occupancy [%]	Distance (STD) [Å]
VAL8-OXT – ARG85-NH2	30.07	5.69 (2.70)
GLU3-O – GLY33-N	22.37	8.30 (4.97)
VAL5-N – VAL31-O	22.12	6.36 (3.49)
SP16-OE2 – ARG85-NH2	20.70	6.12 (3.00)
VAL8-O – ARG25-NE	20.09	6.88 (2.93)
GLN4-O – GLY33-N	19.40	8.05 (4.35)
VAL8-OXT – ARG25-NH2	18.19	7.97 (3.61)
GLU3-N – GLY33-O	17.91	9.14 (5.92)
GLU3-N – ASP35-OD1	16.01	9.35 (6.30)
VAL8-O – ARG25-NH2	15.43	7.67 (3.48)
VAL8-O – ARG85-NH1	14.62	6.62 (2.50)
GLN4-N – GLY33-O	13.00	8.67 (5.12)
ALA7-O – ARG85-NH1	12.89	5.96 (2.21)
VAL8-O – ARG85-NE	12.26	6.28 (2.25)
SP16-OE1 – ARG85-NH2	12.14	6.42 (3.00)
VAL8-OXT – ARG85-NE	11.17	5.82 (2.48)
SP16-OE1 – ARG85-NE	10.83	7.31 (2.79)
ALA7-O – ARG85-NH2	10.74	5.65 (1.81)
GLU3-N – ASP35-OD2	10.73	9.43 (6.13)
VAL8-O – ARG85-NH2	10.58	6.35 (2.55)
SP16-O – VAL31-N	10.32	7.24 (2.27)

TABLE B.12: Percentage of hydrogen-bond occupancies and the distances between the involved heavy atoms for hydrogen bonds between MAGI1 PDZ1 Q85R and EQVpSAV peptide with singly charged (SP1) phosphate, existing in 10 % or more of all simulation frames. In the left column, first the peptide residue and second the protein residue is given.

Appendix C

Supplementary Material for chapter 5

This supporting information belongs to chapter 5.

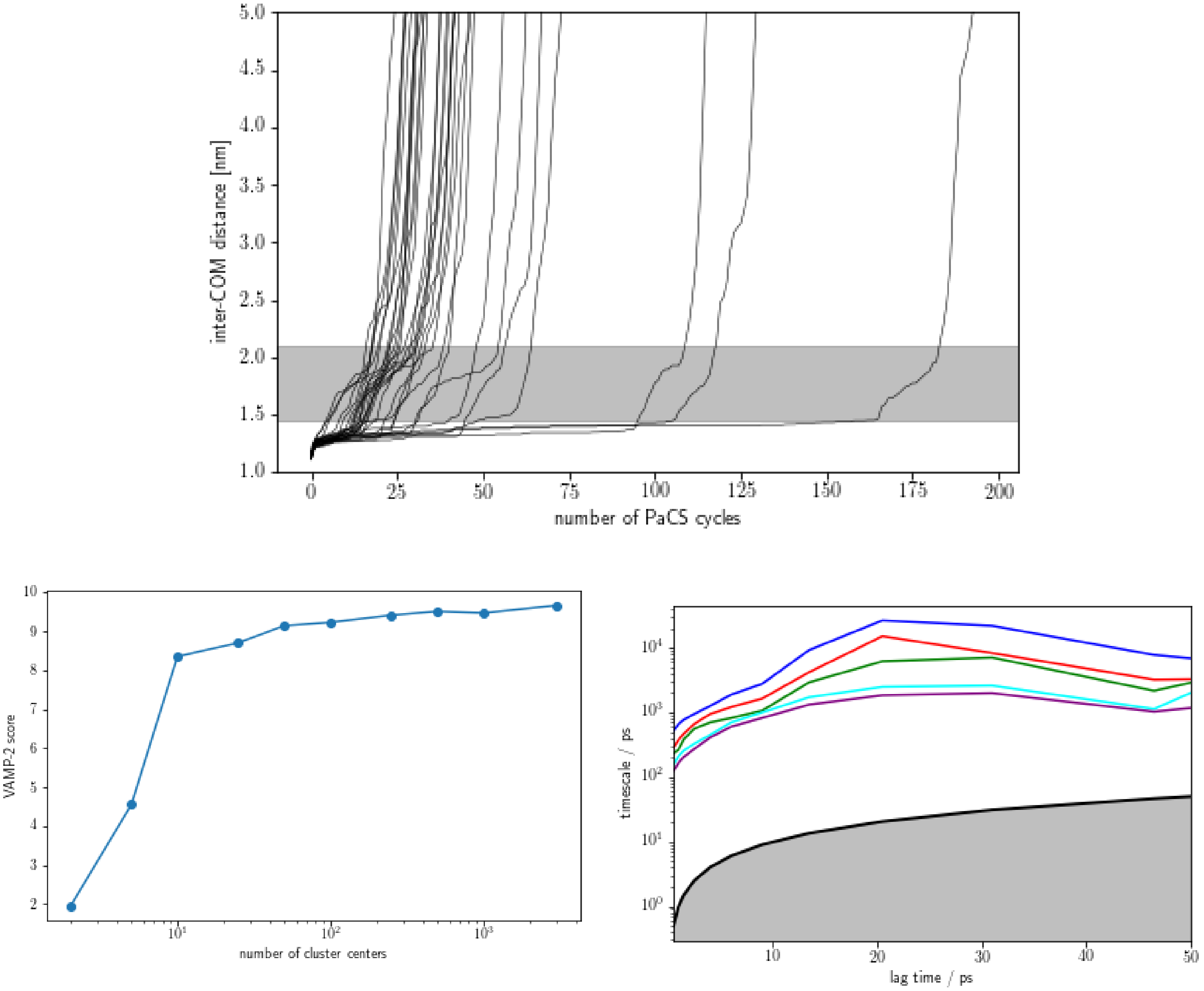


FIGURE C.1: Analysis of the PaCS-MD simulation of the SLKLMTTV peptide binding the mDvl1 PDZ domain. Top: Inter-COM distance as a function of the number of PaCS-MD cycles. The inter-COM distance range is divided in the bound, partially-bound and unbound regions. Bottom left: VAMP-2 scores for as a function of the number of cluster centers. The analysis was repeated five times for every studied number of cluster centers and the standard deviation is shown as light-blue area. Bottom right: Implied MSM time scales as a function of the lag time τ.

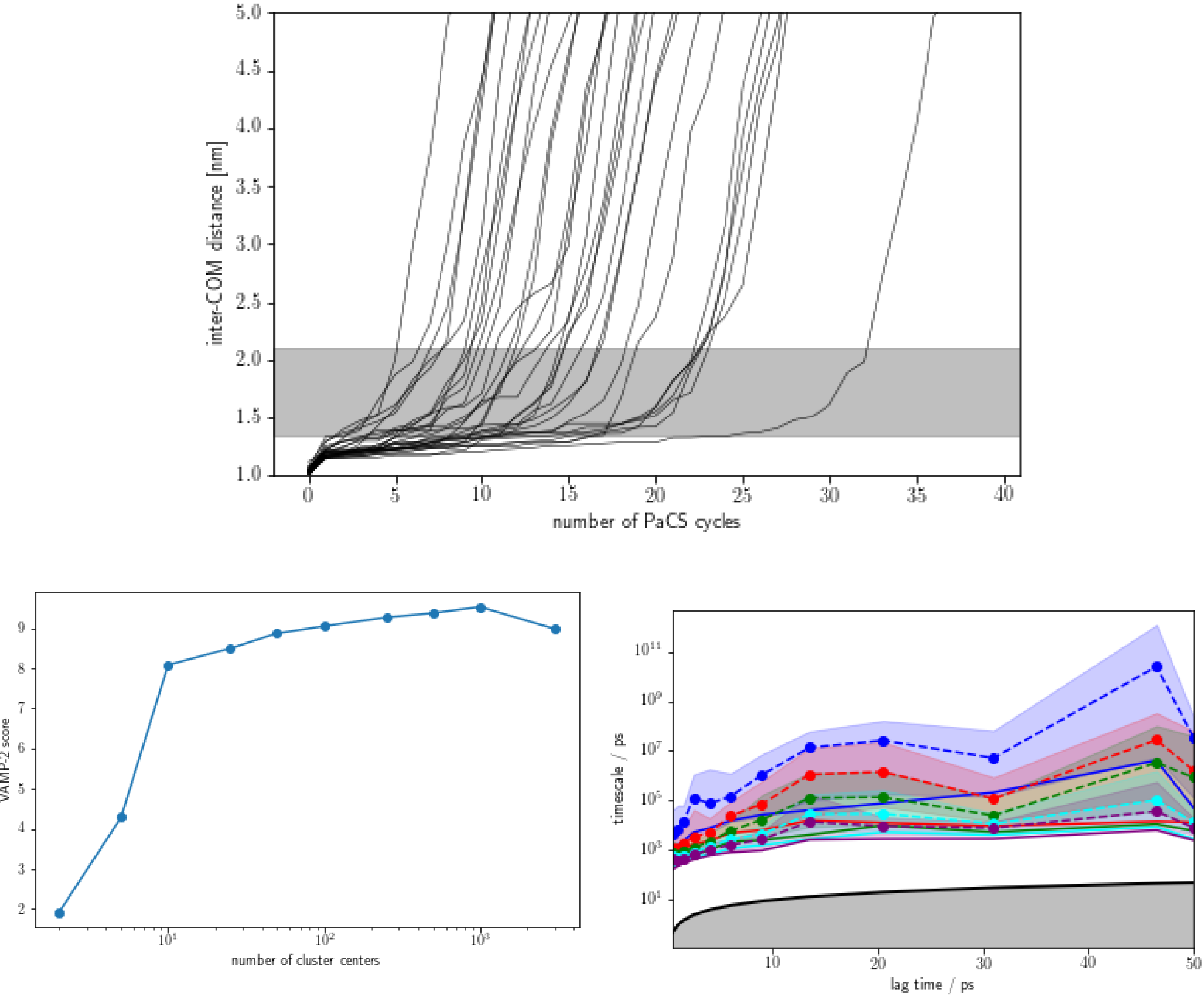


FIGURE C.2: Analysis of the PaCS-MD simulation of the small molecule inhibitor SUZ binding the mDvl1 PDZ domain. Top: Inter-COM distance as a function of the number of PaCS-MD cycles. The inter-COM distance range is divided in the bound, partially-bound and unbound regions. Bottom left: VAMP-2 scores for as a function of the number of cluster centers. The analysis was repeated five times for every studied number of cluster centers and the standard deviation is shown as light-blue area. Bottom right: Implied MSM time scales as a function of the lag time τ.

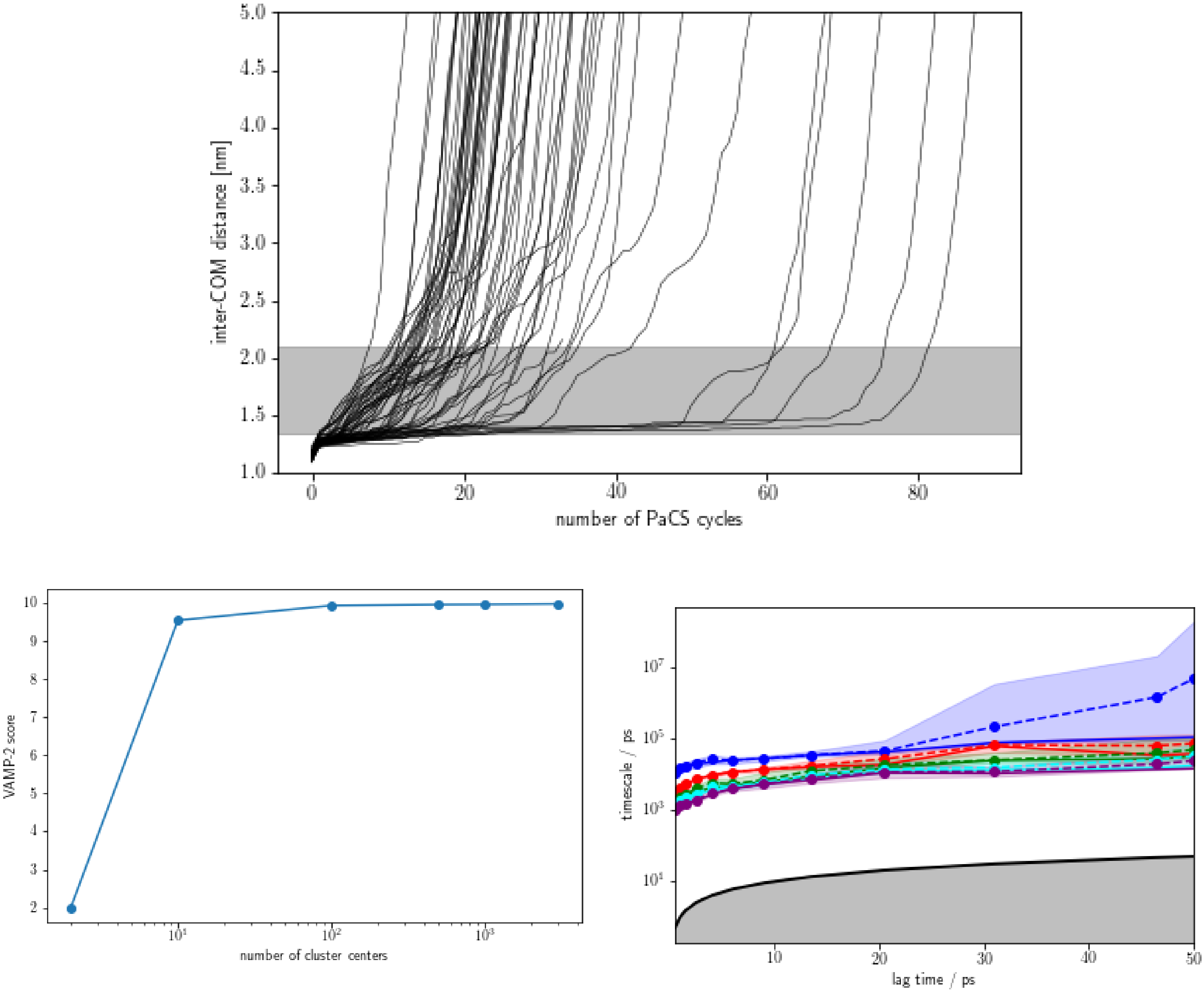


FIGURE C.3: MSM analysis of the PaCS-MD simulation with 34 initial position from a single long plain MD simulation of the SLKLMTTV peptide binding the mDvl1 PDZ domain when the small molecule inhibitor SUZ is competing for the binding pocket. Top: Inter-COM distance as a function of the number of PaCS-MD cycles. The inter-COM distance range is divided in the bound, partially-bound and unbound regions. Bottom left: VAMP-2 scores for as a function of the number of cluster centers. Bottom right: Implied MSM time scales as a function of the lag time τ.

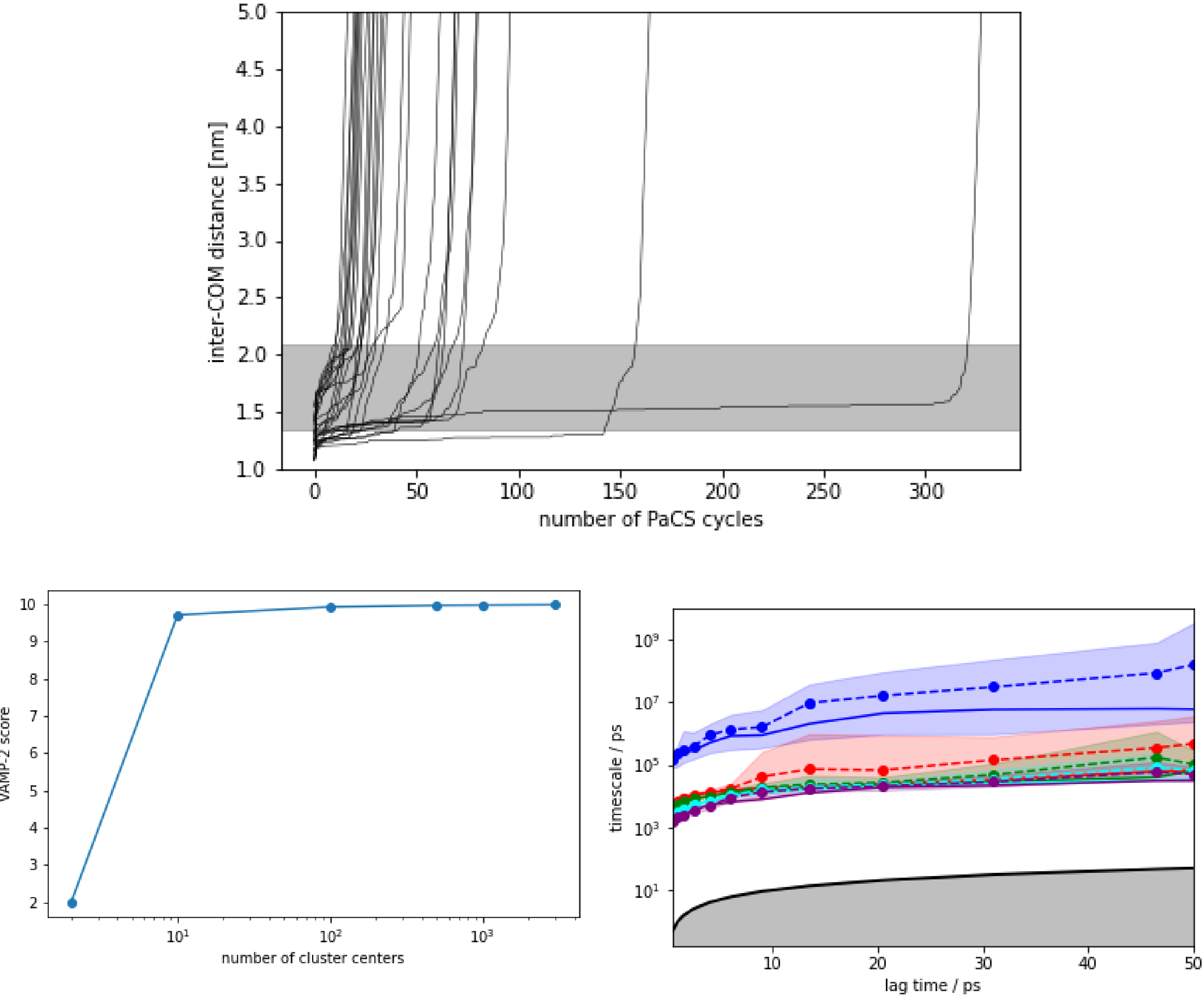


FIGURE C.4: MSM analysis of the PaCS-MD simulation with 27 initial position from the plain MD simulation 1, 2 and 3 of the SLKLMTTV peptide binding the mDvl1 PDZ domain when the small molecule inhibitor SUZ is competing for the binding pocket. Top: Inter-COM distance as a function of the number of PaCS-MD cycles. The inter-COM distance range is divided in the bound, partially-bound and unbound regions. Bottom left: VAMP-2 scores for as a function of the number of cluster centers. Bottom right: Implied MSM time scales as a function of the lag time τ.

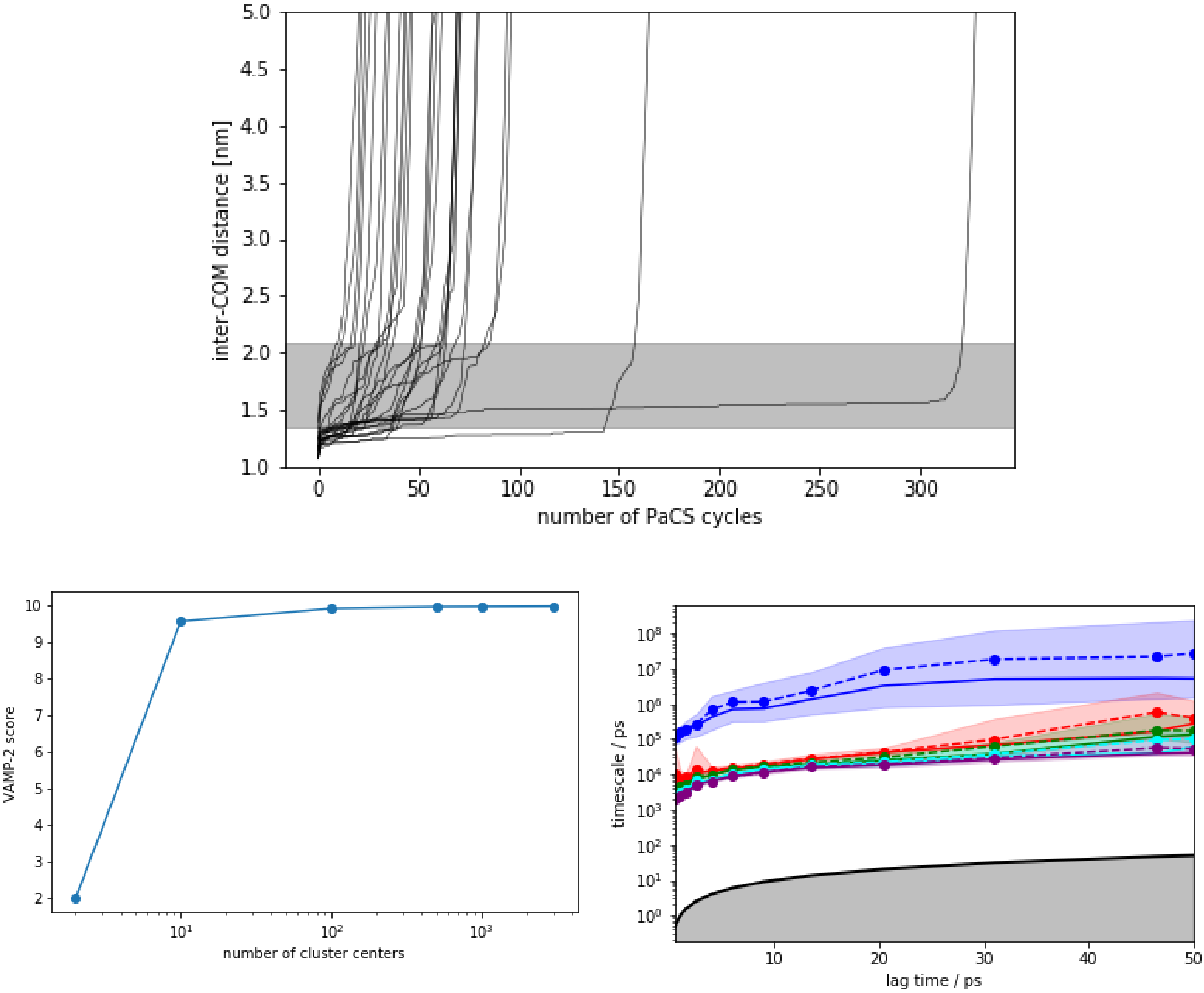


FIGURE C.5: MSM analysis of the PaCS-MD simulation with 27 initial position from the plain MD simulation 1, 2 and 4 of the SLKLMTTV peptide binding the mDvl1 PDZ domain when the small molecule inhibitor SUZ is competing for the binding pocket. Top: Inter-COM distance as a function of the number of PaCS-MD cycles. The inter-COM distance range is divided in the bound, partially-bound and unbound regions. Bottom left: VAMP-2 scores for as a function of the number of cluster centers. Bottom right: Implied MSM time scales as a function of the lag time τ.

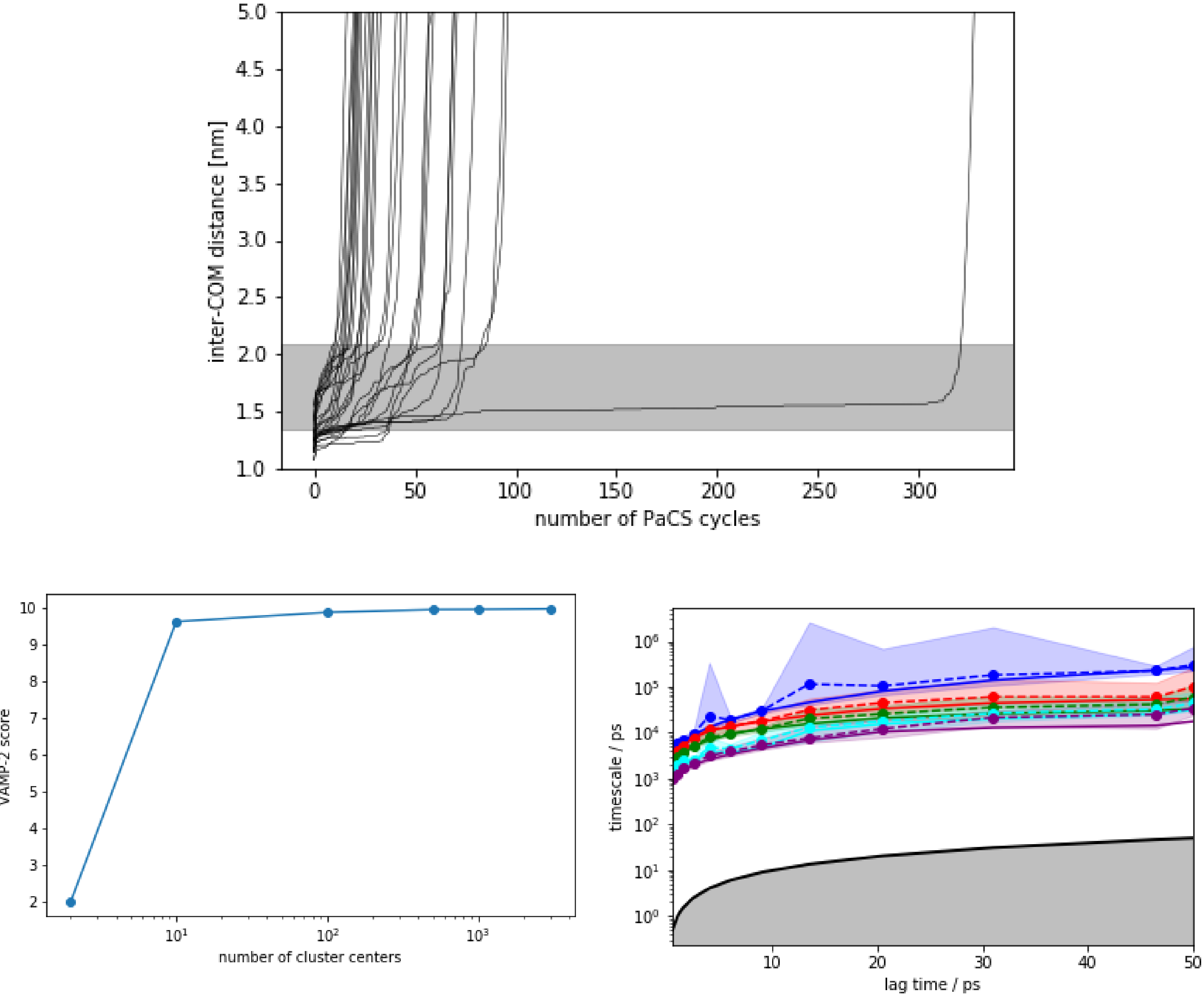


FIGURE C.6: MSM analysis of the PaCS-MD simulation with 27 initial position from the plain MD simulation 1, 3 and 4 of the SLKLMTTV peptide binding the mDvl1 PDZ domain when the small molecule inhibitor SUZ is competing for the binding pocket. Top: Inter-COM distance as a function of the number of PaCS-MD cycles. The inter-COM distance range is divided in the bound, partially-bound and unbound regions. Bottom left: VAMP-2 scores for as a function of the number of cluster centers. Bottom right: Implied MSM time scales as a function of the lag time τ.

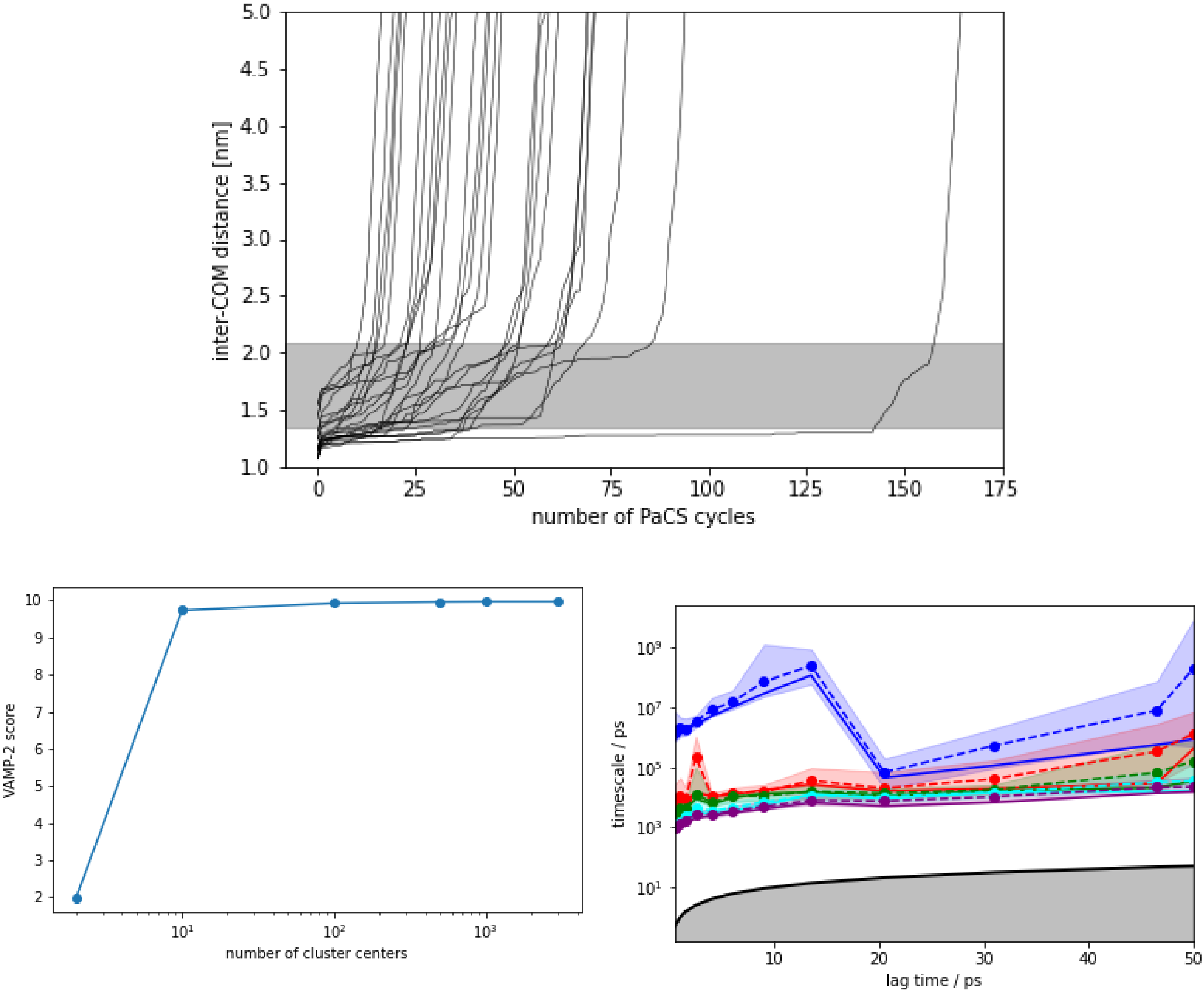


FIGURE C.7: MSM analysis of the PaCS-MD simulation with 27 initial position from the plain MD simulation 2, 3 and 4 of the SLKLMTTV peptide binding the mDvl1 PDZ domain when the small molecule inhibitor SUZ is competing for the binding pocket. Top: Inter-COM distance as a function of the number of PaCS-MD cycles. The inter-COM distance range is divided in the bound, partially-bound and unbound regions. Bottom left: VAMP-2 scores for as a function of the number of cluster centers. Bottom right: Implied MSM time scales as a function of the lag time τ.

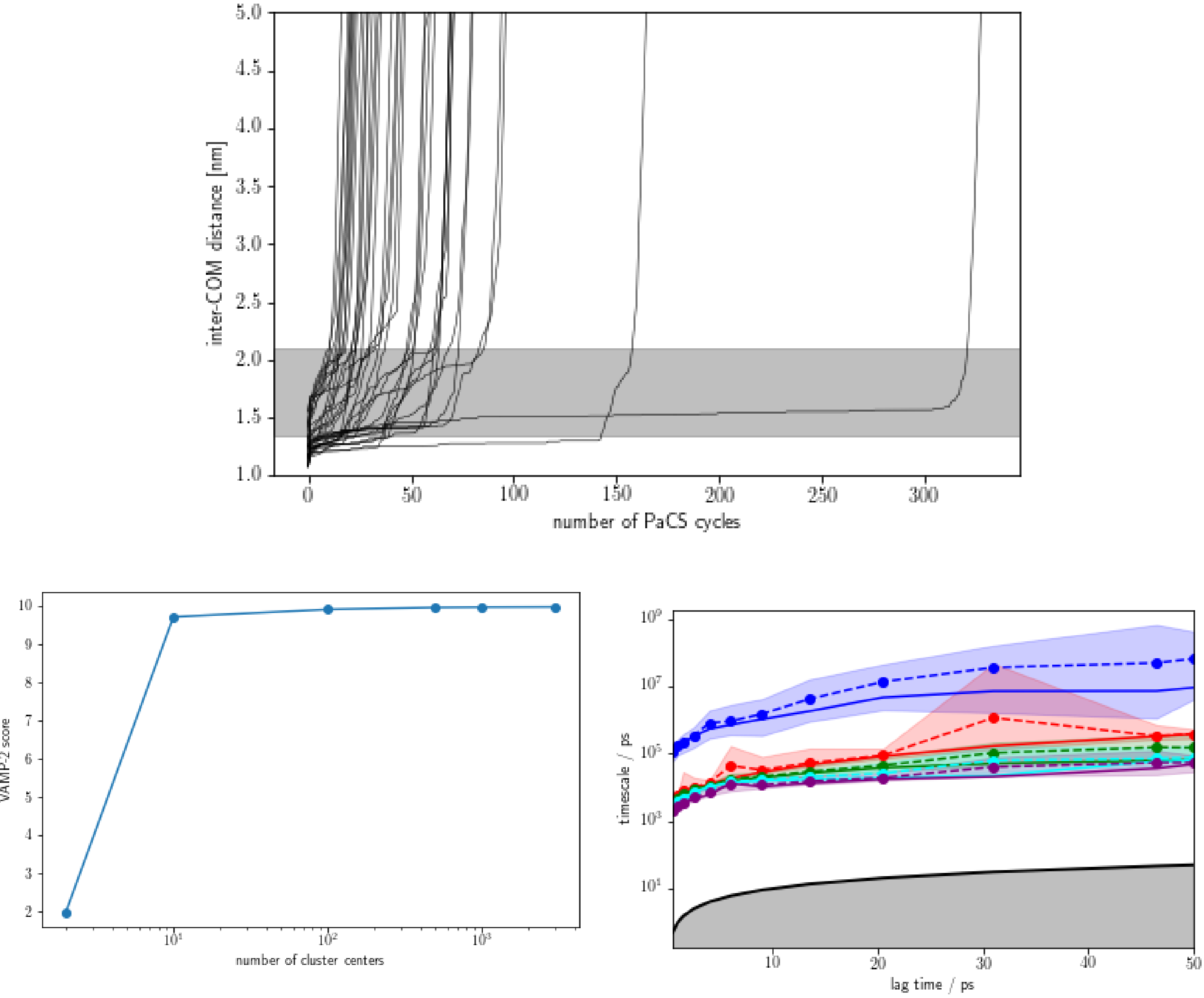


FIGURE C.8: MSM analysis of the PaCS-MD simulation with 36 initial position from the plain MD simulation 1, 2, 3 and 4 of the SLKLMTTV peptide binding the mDvl1 PDZ domain when the small molecule inhibitor SUZ is competing for the binding pocket. Top: Inter-COM distance as a function of the number of PaCS-MD cycles. The inter-COM distance range is divided in the bound, partially-bound and unbound regions. Bottom left: VAMP-2 scores for as a function of the number of cluster centers. Bottom right: Implied MSM time scales as a function of the lag time τ.

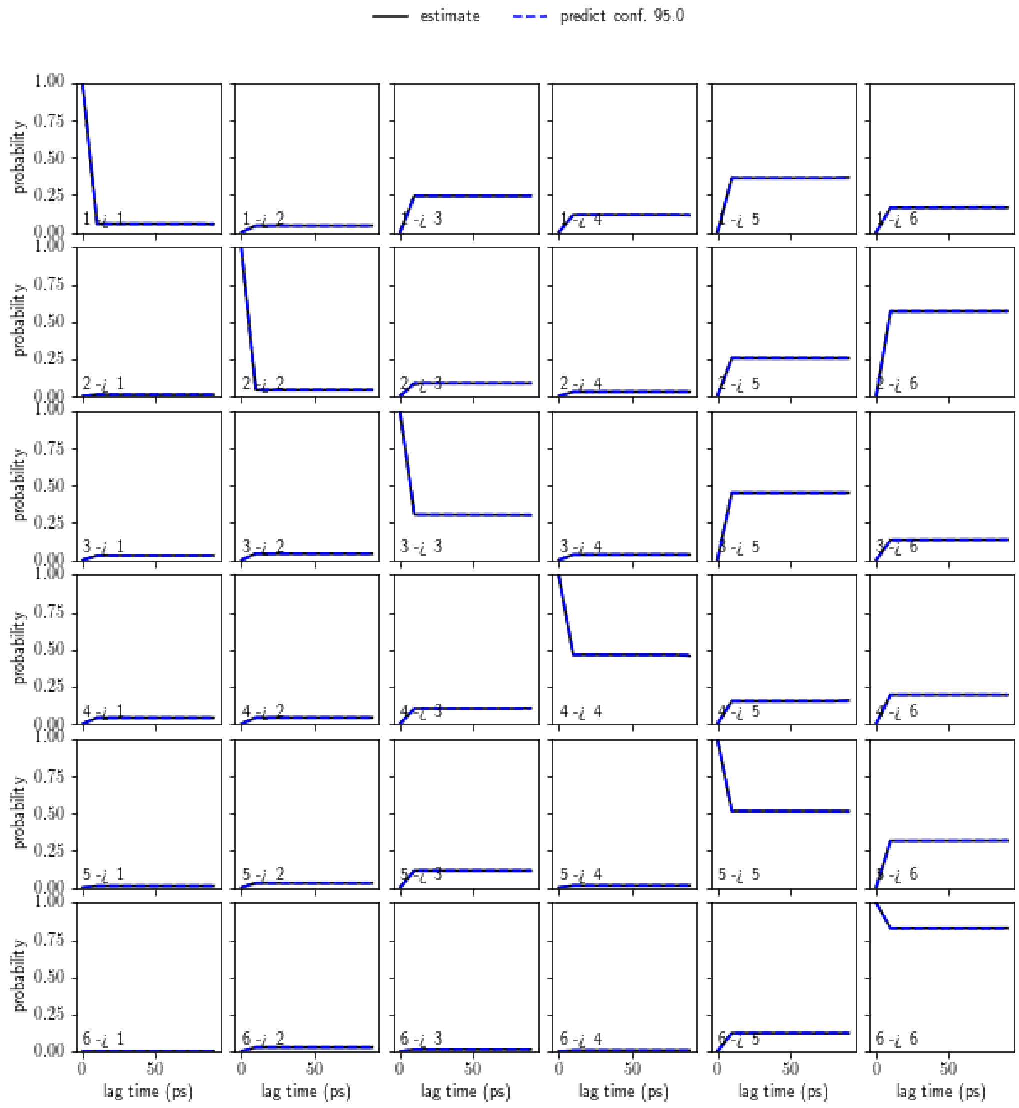


FIGURE C.9: Chapman-Kolmogorov test for the MSM built from the 34 initial positions extracted from the long plain MD simulation labeled 0. It is clear, that the MSM fulfills Markovianity.

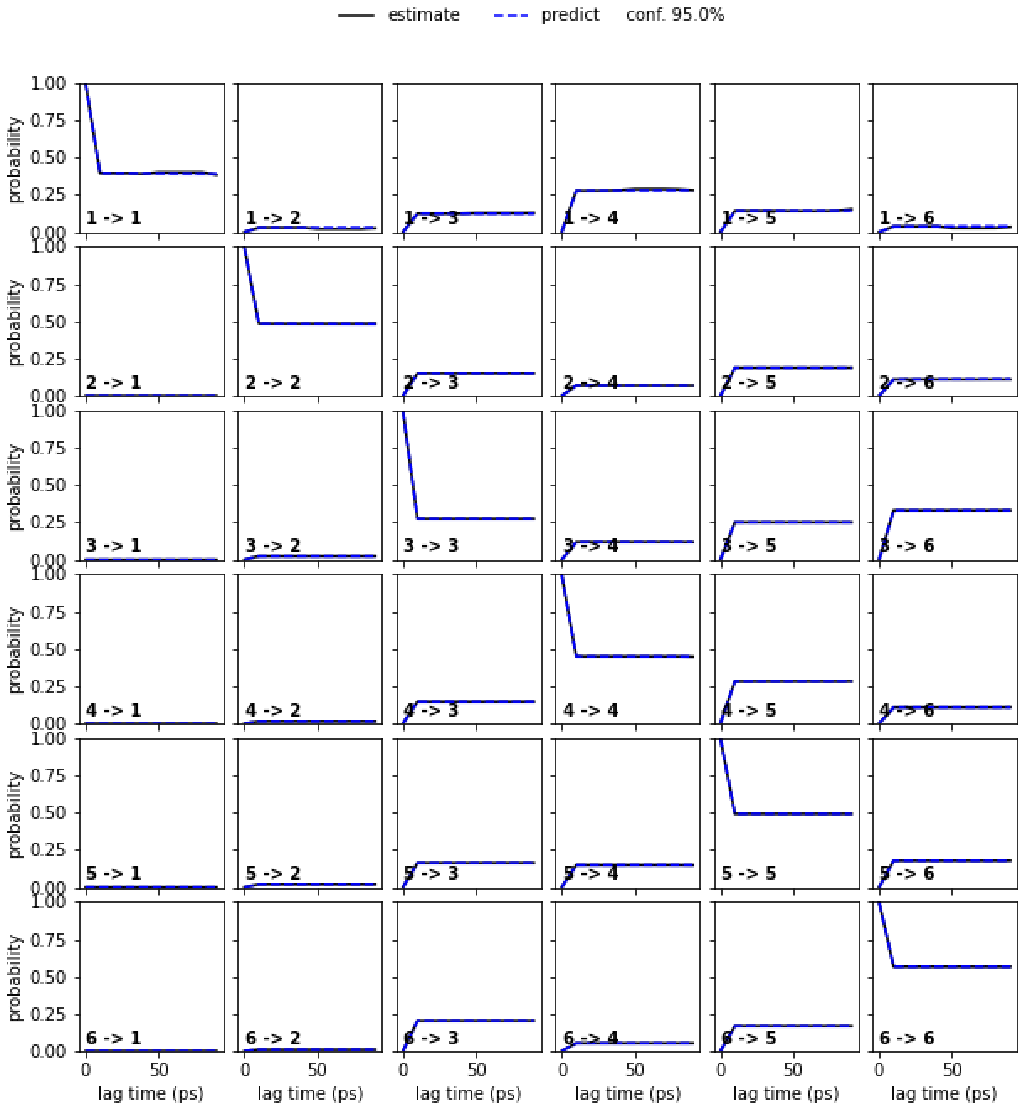


FIGURE C.10: Chapman-Kolmogorov test for the MSM built from the 27 initial positions extracted from the three plain MD simulation labeled 1, 2 and 3. It is clear, that the MSM fulfills Markovianity.

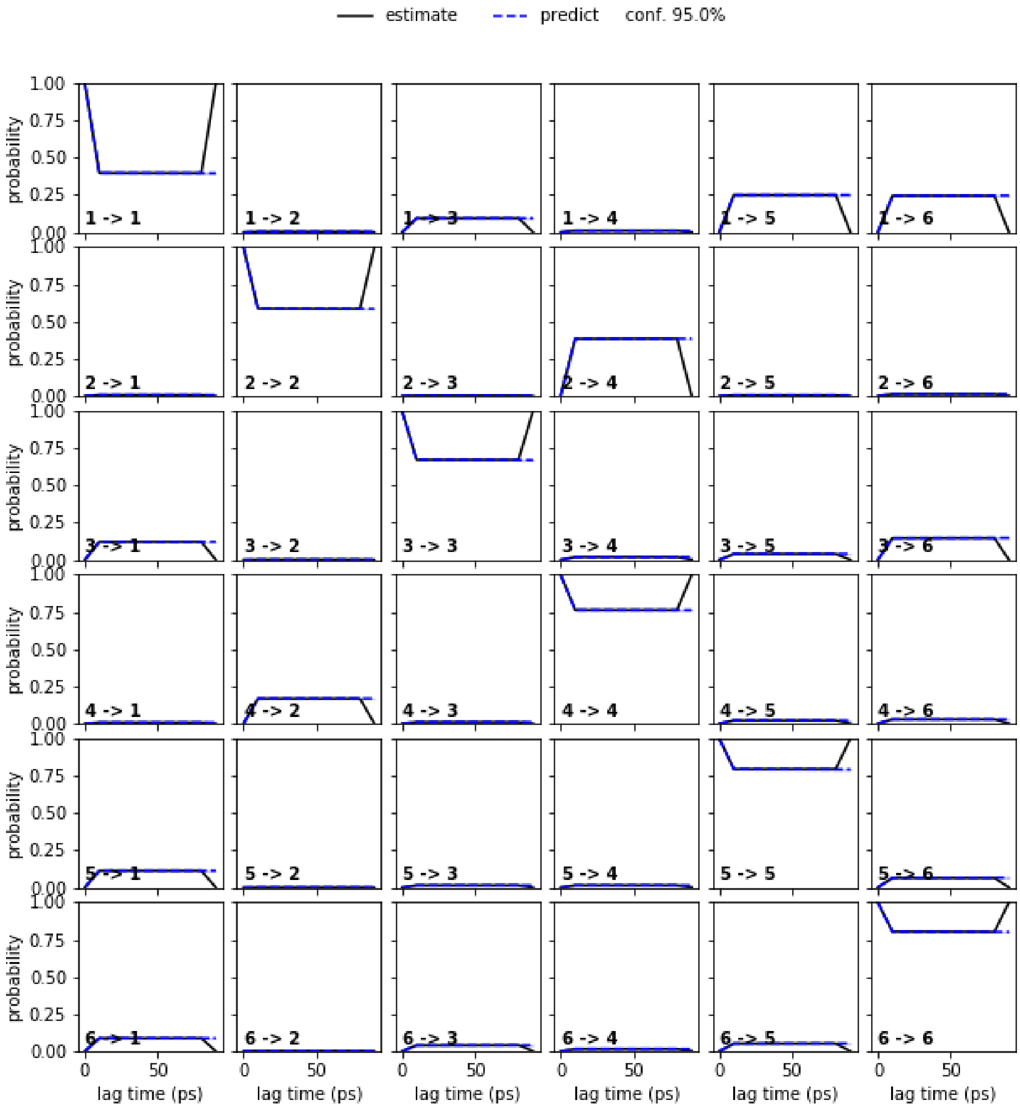


FIGURE C.11: Chapman-Kolmogorov test for the MSM built from the 27 initial positions extracted from the three plain MD simulation labeled 1, 2 and 4. It is clear, that the MSM fulfills Markovianity.

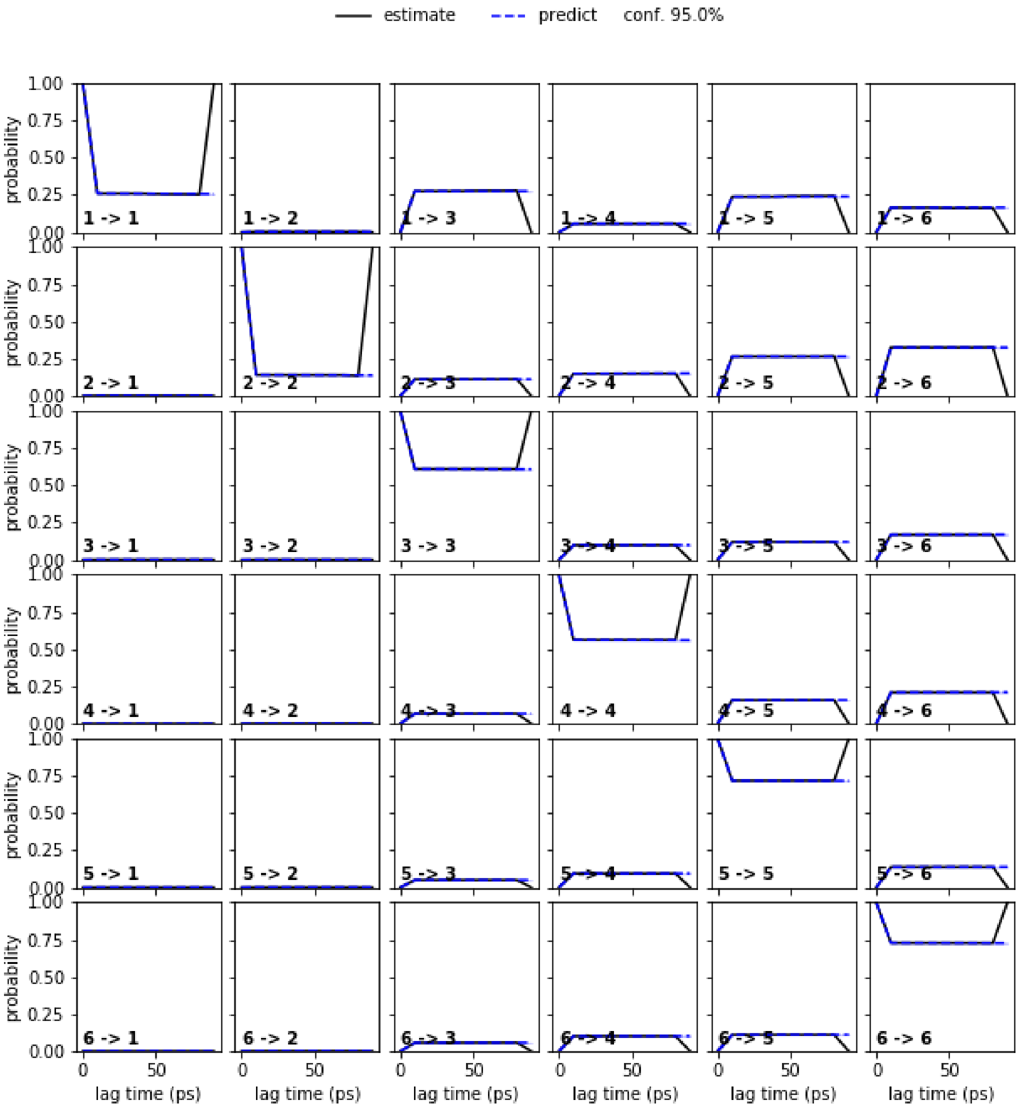


FIGURE C.12: Chapman-Kolmogorov test for the MSM built from the 27 initial positions extracted from the three plain MD simulation labeled 1, 3 and 4. It is clear, that the MSM fulfills Markovianity.

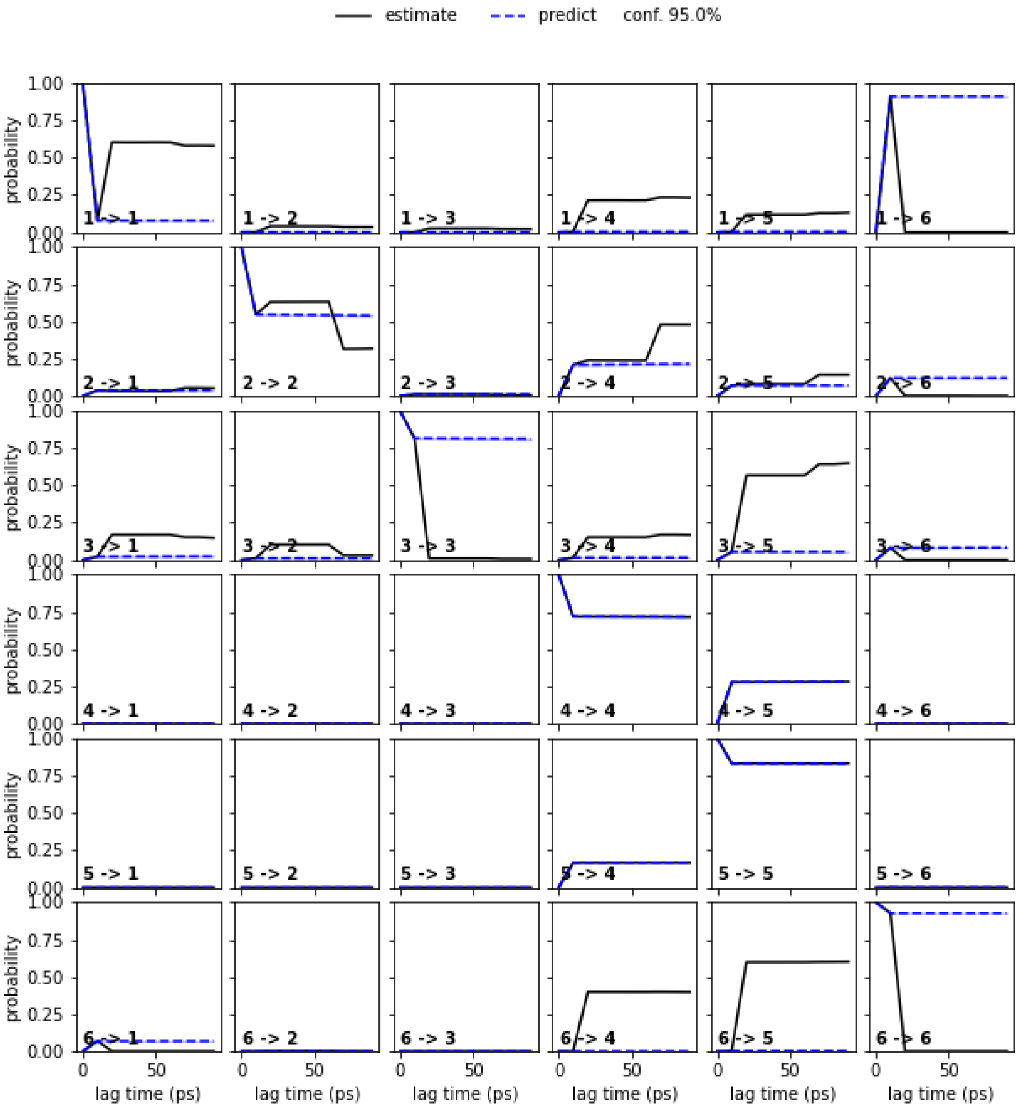


FIGURE C.13: Chapman-Kolmogorov test for the MSM built from the 27 initial positions extracted from the three plain MD simulation labeled 2, 3 and 4. It is clear, that the MSM fulfills Markovianity.

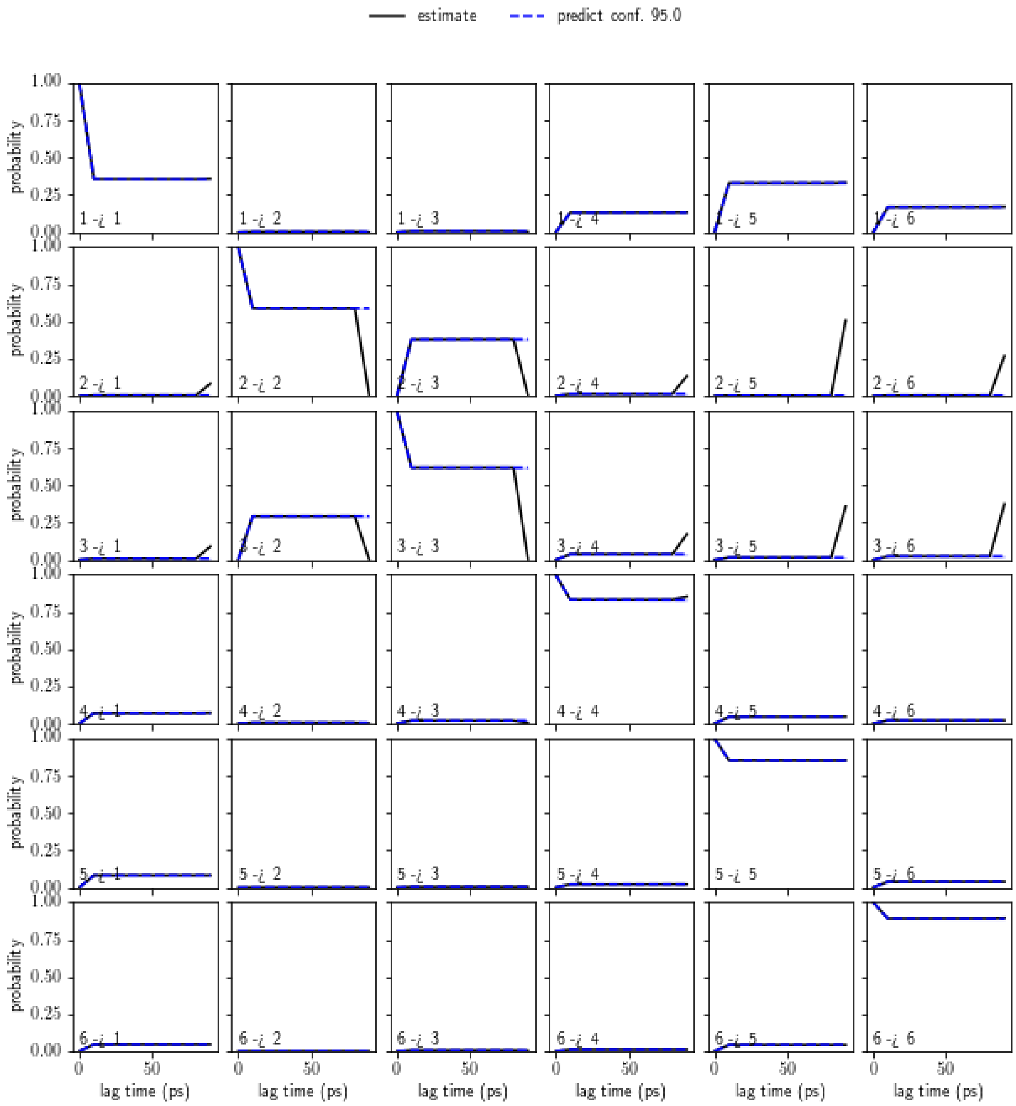


FIGURE C.14: Chapman-Kolmogorov test for the MSM built from the 36 initial positions extracted from the four plain MD simulation labeled 1, 2, 3 and 4. It is clear, that the MSM fulfills Markovianity.

Bibliography

[1] K. Vanommeslaeghe et al. "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields". In: *Journal of Computational Chemistry* 31.4 (2010), pp. 671–690. DOI: `https://doi.org/10.1002/jcc.21367`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.21367`.

[2] Wenbo Yu et al. "Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations". In: *Journal of Computational Chemistry* 33.31 (2012), pp. 2451–2468. DOI: `https://doi.org/10.1002/jcc.23067`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.23067`.

[3] B. R. Brooks et al. "CHARMM: The biomolecular simulation program". In: *Journal of Computational Chemistry* 30.10 (2009), pp. 1545–1614. DOI: `https://doi.org/10.1002/jcc.21287`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.21287`.

[4] Mark James Abraham et al. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". In: *SoftwareX* 1-2 (2015), pp. 19 –25. ISSN: 2352-7110. DOI: `10.1016/j.softx.2015.06.001`. URL: `http://www.sciencedirect.com/science/article/pii/S2352711015000059`.

[5] Chr. Møller and M. S. Plesset. "Note on an Approximation Treatment for Many-Electron Systems". In: *Phys. Rev.* 46 (7 Oct. 1934), pp. 618–622. DOI: `10.1103/PhysRev.46.618`. URL: `https://link.aps.org/doi/10.1103/PhysRev.46.618`.

[6] Helen M. Berman et al. "The Protein Data Bank". In: *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 235–242. ISSN: 0305-1048. DOI: `10.1093/nar/28.1.235`. URL: `https://doi.org/10.1093/nar/28.1.235`.

[7] Helen Berman, Kim Henrick, and Haruki Nakamura. "Announcing the worldwide Protein Data Bank". In: *Nature Structural & Molecular Biology* 10.12 (Dec. 2003), pp. 980–980. ISSN: 1545-9985. DOI: `10.1038/nsb1203-980`. URL: `https://doi.org/10.1038/nsb1203-980`.

[8] Tom Darden, Darrin York, and Lee Pedersen. "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems". In: *The Journal of Chemical Physics* 98.12 (1993), pp. 10089–10092. DOI: `10.1063/1.464397`. URL: `https://doi.org/10.1063/1.464397`.

[9] Ulrich Essmann et al. "A smooth particle mesh Ewald method". In: *The Journal of Chemical Physics* 103.19 (1995), pp. 8577–8593. DOI: `10.1063/1.470117`. URL: `https://doi.org/10.1063/1.470117`.

[10] H. J. C. Berendsen et al. “Interaction Models for Water in Relation to Protein Hydration”. In: *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*. Ed. by Bernard Pullman. Dordrecht: Springer Netherlands, 1981, pp. 331–342. ISBN: 978-94-015-7658-1. DOI: `10.1007/978-94-015-7658-1_21`. URL: `https://doi.org/10.1007/978-94-015-7658-1_21`.

[11] *PubChem Compound Summary for CID 1548887, Sulindac*. Sept. 2021. URL: `https://pubchem.ncbi.nlm.nih.gov/compound/Sulindac`.

[12] L. Molgedey and H. G. Schuster. “Separation of a mixture of independent signals using time delayed correlations”. In: *Phys. Rev. Lett.* 72 (23 June 1994), pp. 3634–3637. DOI: `10.1103/PhysRevLett.72.3634`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.72.3634`.

[13] Guillermo Pérez-Hernández et al. “Identification of slow molecular order parameters for Markov model construction”. In: *The Journal of Chemical Physics* 139.1 (2013), p. 015102. DOI: `10.1063/1.4811489`. URL: `https://doi.org/10.1063/1.4811489`.

[14] Christian R. Schwantes and Vijay S. Pande. “Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9”. In: *Journal of Chemical Theory and Computation* 9.4 (2013). PMID: 23750122, pp. 2000–2009. DOI: `10.1021/ct300878a`. URL: `https://doi.org/10.1021/ct300878a`.

[15] Frank Noé and Cecilia Clementi. “Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation”. In: *Journal of Chemical Theory and Computation* 11.10 (2015), pp. 5002–5011. DOI: `10.1021/acs.jctc.5b00553`. URL: `https://doi.org/10.1021/acs.jctc.5b00553`.

[16] Philipp Metzner, Christof Schütte, and Eric Vanden-Eijnden. “Transition Path Theory for Markov Jump Processes”. In: *Multiscale Modeling & Simulation* 7.3 (2009), pp. 1192–1219. DOI: `10.1137/070699500`. URL: `https://doi.org/10.1137/070699500`.

[17] Ron Milo. “What is the total number of protein molecules per cell volume? A call to rethink some published values”. In: *BioEssays* 35.12 (2013), pp. 1050–1055. DOI: `https://doi.org/10.1002/bies.201300066`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/bies.201300066`.

[18] Francisco Feijó Delgado et al. “Intracellular Water Exchange for Measuring the Dry Mass, Water Mass and Changes in Chemical Composition of Living Cells”. In: *PLOS ONE* 8.7 (July 2013), pp. 1–11. DOI: `10.1371/journal.pone.0067590`. URL: `https://doi.org/10.1371/journal.pone.0067590`.

[19] S Jones and J M Thornton. “Principles of protein-protein interactions”. In: *Proceedings of the National Academy of Sciences* 93.1 (1996), pp. 13–20. ISSN: 0027-8424. DOI: `10.1073/pnas.93.1.13`. URL: `https://www.pnas.org/content/93/1/13`.

[20] Joseph A. Marsh and Sarah A. Teichmann. “Structure, Dynamics, Assembly, and Evolution of Protein Complexes”. In: *Annual Review of Biochemistry* 84.1 (2015). PMID: 25494300, pp. 551–575. DOI: `10.1146/annurev-biochem-060614-034142`. URL: `https://doi.org/10.1146/annurev-biochem-060614-034142`.

[21] John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* (July 2021). ISSN: 1476-4687. DOI: `10.1038/s41586-021-03819-2`. URL: `https://doi.org/10.1038/s41586-021-03819-2`.

[22] Michael Shirts and Vijay S. Pande. "Screen Savers of the World Unite!" In: *Science* 290.5498 (2000), pp. 1903–1904. ISSN: 0036-8075. DOI: `10.1126/science.290.5498.1903`. URL: `https://science.sciencemag.org/content/290/5498/1903`.

[23] Adam L. Beberg et al. "Folding@home: Lessons from eight years of volunteer distributed computing". In: *2009 IEEE International Symposium on Parallel Distributed Processing*. 2009, pp. 1–8. DOI: `10.1109/IPDPS.2009.5160922`.

[24] T. Giorgino, I. Buch, and G. De Fabritiis. "Visualizing the Induced Binding of SH2-Phosphopeptide". In: *Journal of Chemical Theory and Computation* 8.4 (2012). PMID: 26596736, pp. 1171–1175. DOI: `10.1021/ct300003f`. URL: `https://doi.org/10.1021/ct300003f`.

[25] Nicolas Blöchliger, Min Xu, and Amedeo Caflisch. "Peptide Binding to a PDZ Domain by Electrostatic Steering via Nonnative Salt Bridges". In: *Biophysical Journal* 108.9 (2015), pp. 2362–2370. ISSN: 0006-3495. DOI: `doi.org/10.1016/j.bpj.2015.03.038`. URL: `https://www.sciencedirect.com/science/article/pii/S0006349515002994`.

[26] James C. Gumbart, Benoît Roux, and Christophe Chipot. "Standard binding free energies from computer simulations: What is the best strategy?" eng. In: *Journal of chemical theory and computation* 9.1 (Jan. 2013). PMC3685508[pmcid], pp. 794–802. ISSN: 1549-9626. DOI: `10.1021/ct3008099`. URL: `https://doi.org/10.1021/ct3008099`.

[27] Duy Phuoc Tran and Akio Kitao. "Dissociation Process of a MDM2/p53 Complex Investigated by Parallel Cascade Selection Molecular Dynamics and the Markov State Model". In: *The Journal of Physical Chemistry B* 123.11 (2019). PMID: 30645121, pp. 2469–2478. DOI: `10.1021/acs.jpcb.8b10309`. URL: `https://doi.org/10.1021/acs.jpcb.8b10309`.

[28] Jayvee R. Abella et al. "Markov state modeling reveals alternative unbinding pathways for peptide–MHC complexes". In: *Proceedings of the National Academy of Sciences* 117.48 (2020), pp. 30610–30618. ISSN: 0027-8424. DOI: `10.1073/pnas.2007246117`. URL: `https://www.pnas.org/content/117/48/30610`.

[29] Alexsandra B. Espejo et al. "PRMT5 C-terminal phosphorylation modulates a 14-3-3/PDZ interaction switch". In: *Journal of Biological Chemistry* 292.6 (2017), pp. 2255–2265. ISSN: 1083351X. DOI: `10.1074/jbc.M116.760330`.

[30] Stephan Tetenborg et al. "Phosphorylation of Connexin36 near the C-terminus switches binding affinities for PDZ-domain and 14–3–3 proteins in vitro". In: *Scientific Reports* 10.1 (Oct. 2020), p. 18378. ISSN: 2045-2322. DOI: `10.1038/s41598-020-75375-0`. URL: `https://doi.org/10.1038/s41598-020-75375-0`.

[31] Gergo Gogl et al. "Dual Specificity PDZ- and 14-3-3-Binding Motifs: A Structural and Interactomics Study". In: *Structure* 28.7 (2020), 747–759.e3. ISSN: 0969-2126. DOI: `doi.org/10.1016/j.str.2020.03.010`. URL: `https://www.sciencedirect.com/science/article/pii/S0969212620300927`.

[32] Nicolas Künzel and Volkhard Helms. "How phosphorylation of peptides affects their interaction with 14-3-3η domains". In: *Proteins: Structure, Function, and Bioinformatics* n/a.n/a (). DOI: `https://doi.org/10.1002/prot.26224`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.26224`.

[33] Simone Trautmann et al. "Deep Proteomic Insights into the Individual Short-Term Pellicle Formation on Enamel-An In Situ Pilot Study". In: *PROTEOMICS – Clinical Applications* 14.3 (2020), p. 1900090. DOI: `https://doi.org/10.1002/prca.201900090`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/prca.201900090`.

[34] Mona L. Knapp et al. "Alternative splicing switches STIM1 targeting to specialized membrane contact sites and modifies SOCE". In: *bioRxiv* (2020). DOI: `10.1101/2020.03.25.005199`. URL: `https://www.biorxiv.org/content/early/2020/03/25/2020.03.25.005199`.

[35] Irfan ur Rauf Tak et al. "Chapter 1 - Posttranslational Modifications of Proteins and Their Role in Biological Processes and Associated Diseases". In: *Protein Modificomics*. Ed. by Tanveer Ali Dar and Laishram Rajendrakumar Singh. Academic Press, 2019, pp. 1–35. ISBN: 978-0-12-811913-6. DOI: `https://doi.org/10.1016/B978-0-12-811913-6.00001-1`. URL: `https://www.sciencedirect.com/science/article/pii/B9780128119136000011`.

[36] Sudhakaran Prabakaran et al. "Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding". In: *WIREs Systems Biology and Medicine* 4.6 (2012), pp. 565–583. DOI: `https://doi.org/10.1002/wsbm.1185`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/wsbm.1185`.

[37] International Human Genome Sequencing Consortium. "Finishing the euchromatic sequence of the human genome". In: *Nature* 431.7011 (Oct. 2004), pp. 931–945. ISSN: 1476-4687. DOI: `10.1038/nature03001`. URL: `https://doi.org/10.1038/nature03001`.

[38] Ole Nørregaard Jensen. "Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry". In: *Current Opinion in Chemical Biology* 8.1 (2004), pp. 33–41. ISSN: 1367-5931. DOI: `https://doi.org/10.1016/j.cbpa.2003.12.009`. URL: `https://www.sciencedirect.com/science/article/pii/S1367593103001789`.

[39] Yonathan Lissanu Deribe, Tony Pawson, and Ivan Dikic. "Post-translational modifications in signal integration". In: *Nature Structural & Molecular Biology* 17.6 (June 2010), pp. 666–672. ISSN: 1545-9985. DOI: `10.1038/nsmb.1842`. URL: `https://doi.org/10.1038/nsmb.1842`.

[40] Raj B Parekh and Christian Rohlff. "Post-translational modification of proteins and the discovery of new medicine". In: *Current Opinion in Biotechnology* 8.6 (1997), pp. 718–723. ISSN: 0958-1669. DOI: `https://doi.org/10.1016/S0958-1669(97)80126-7`. URL: `https://www.sciencedirect.com/science/article/pii/S0958166997801267`.

[41] Ann M. Bode and Zigang Dong. "Post-translational modification of p53 in tumorigenesis". In: *Nature Reviews Cancer* 4.10 (Oct. 2004), pp. 793–805. ISSN: 1474-1768. DOI: `10.1038/nrc1455`. URL: `https://doi.org/10.1038/nrc1455`.

[42] Karl E. Krueger and Sudhir Srivastava. "Posttranslational Protein Modifications: Current Implications for Cancer Detection, Prevention, and Therapeutics *". In: *Molecular & Cellular Proteomics* 5.10 (Oct. 2006), pp. 1799–1810. ISSN: 1535-9476. DOI: `10.1074/mcp.R600009-MCP200`. URL: `https://doi.org/10.1074/mcp.R600009-MCP200`.

[43] Travis Van der Steen, Donald J. Tindall, and Haojie Huang. "Posttranslational Modification of the Androgen Receptor in Prostate Cancer". In: *International Journal of Molecular Sciences* 14.7 (2013), pp. 14833–14859. ISSN: 1422-0067. DOI: `10.3390/ijms140714833`. URL: `https://www.mdpi.com/1422-0067/14/7/14833`.

[44] Zhi-Jian Han et al. "The post-translational modification, SUMOylation, and cancer (Review)". In: *Int J Oncol* 52.4 (Apr. 2018), pp. 1081–1094. DOI: `10.3892/ijo.2018.4280`. URL: `https://doi.org/10.3892/ijo.2018.4280`.

[45] Kai-Yao Huang et al. "dbPTM in 2019: exploring disease association and cross-talk of post-translational modifications". In: *Nucleic Acids Research* 47.D1 (Nov. 2018), pp. D298–D308. ISSN: 0305-1048. DOI: `10.1093/nar/gky1074`. URL: `https://doi.org/10.1093/nar/gky1074`.

[46] Prathibha R. Gajjala et al. "Emerging role of post-translational modifications in chronic kidney disease and cardiovascular disease". In: *Nephrology Dialysis Transplantation* 30.11 (Apr. 2015), pp. 1814–1824. ISSN: 0931-0509. DOI: `10.1093/ndt/gfv048`. URL: `https://doi.org/10.1093/ndt/gfv048`.

[47] Suruchi Aggarwal et al. "Post-translational Modification Crosstalk and Hotspots in Sirtuin Interactors Implicated in Cardiovascular Diseases". In: *Frontiers in Genetics* 11 (2020), p. 356. ISSN: 1664-8021. DOI: `10.3389/fgene.2020.00356`. URL: `https://www.frontiersin.org/article/10.3389/fgene.2020.00356`.

[48] Jinlin Liu, Li Zhong, and Rui Guo. "The Role of Posttranslational Modification and Mitochondrial Quality Control in Cardiovascular Diseases". In: *Oxidative Medicine and Cellular Longevity* 2021 (Feb. 2021), p. 6635836. ISSN: 1942-0900. DOI: `10.1155/2021/6635836`. URL: `https://doi.org/10.1155/2021/6635836`.

[49] Xue Liang Du et al. "Hyperglycemia inhibits endothelial nitric oxide synthase activity by posttranslational modification at the Akt site". In: *The Journal of Clinical Investigation* 108.9 (Nov. 2001), pp. 1341–1348. DOI: `10.1172/JCI11235`. URL: `https://www.jci.org/articles/view/11235`.

[50] Wagner B. Dias and Gerald W. Hart. "O-GlcNAc modification in diabetes and Alzheimer's disease". In: *Mol. BioSyst.* 3 (11 2007), pp. 766–772. DOI: `10.1039/B704905F`. URL: `http://dx.doi.org/10.1039/B704905F`.

[51] Jessica L. Dunne et al. "Posttranslational Modifications of Proteins in Type 1 Diabetes: The Next Step in Finding the Cure?" In: *Diabetes* 61.8 (2012), pp. 1907–1914. ISSN: 0012-1797. DOI: `10.2337/db11-1675`. URL: `https://diabetes.diabetesjournals.org/content/61/8/1907`.

[52] Fritz A. Lipmann and P.A. Levene. "SERINEPHOSPHORIC ACID OBTAINED ON HYDROLYSIS OF VITELLINIC ACID". In: *Journal of Biological Chemistry* 98.1 (1932), pp. 109–114. ISSN: 0021-9258. DOI: `https://doi.org/10.1016/S0021-9258(18)76142-5`. URL: `https://www.sciencedirect.com/science/article/pii/S0021925818761425`.

[53] George Burnett and Eugene P. Kennedy. “The enzymatic phosphorylation of proteins”. In: *The Journal of biological chemistry* 211.2 (Dec. 1954), 969—980. ISSN: 0021-9258.

[54] George A. Khoury, Richard C. Baliban, and Christodoulos A. Floudas. “Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database”. In: *Scientific Reports* 1.1 (Sept. 2011), p. 90. ISSN: 2045-2322. DOI: `10.1038/srep00090`. URL: `https://doi.org/10.1038/srep00090`.

[55] Lorenzo A Pinna and Maria Ruzzene. “How do protein kinases recognize their substrates?” In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1314.3 (1996), pp. 191–225. ISSN: 0167-4889. DOI: `https://doi.org/10.1016/S0167-4889(96)00083-3`. URL: `https://www.sciencedirect.com/science/article/pii/S0167488996000833`.

[56] Philip Cohen. “The regulation of protein function by multisite phosphorylation - a 25 year update”. In: *Trends in Biochemical Sciences* 25.12 (2000), pp. 596–601. ISSN: 0968-0004. DOI: `https://doi.org/10.1016/S0968-0004(00)01712-6`. URL: `https://www.sciencedirect.com/science/article/pii/S0968000400017126`.

[57] Jeffrey A. Ubersax and James E. Ferrell Jr. “Mechanisms of specificity in protein phosphorylation”. In: *Nature Reviews Molecular Cell Biology* 8.7 (July 2007), pp. 530–541. ISSN: 1471-0080. DOI: `10.1038/nrm2203`. URL: `https://doi.org/10.1038/nrm2203`.

[58] Abderahmane Derouiche, Charlotte Cousin, and Ivan Mijakovic. “Protein phosphorylation from the perspective of systems biology”. In: *Current Opinion in Biotechnology* 23.4 (2012). Nanobiotechnology • Systems biology, pp. 585–590. ISSN: 0958-1669. DOI: `https://doi.org/10.1016/j.copbio.2011.11.008`. URL: `https://www.sciencedirect.com/science/article/pii/S0958166911007191`.

[59] Hafumi Nishi, Kosuke Hashimoto, and Anna R. Panchenko. “Phosphorylation in Protein-Protein Binding: Effect on Stability and Function”. In: *Structure* 19.12 (2011), pp. 1807–1815. ISSN: 0969-2126. DOI: `https://doi.org/10.1016/j.str.2011.09.021`. URL: `https://www.sciencedirect.com/science/article/pii/S0969212611003340`.

[60] Christoph Antz et al. “Control of K+ channel gating by protein phosphorylation: structural switches of the inactivation gate”. In: *Nature Structural Biology* 6.2 (Feb. 1999), pp. 146–150. ISSN: 1545-9985. DOI: `10.1038/5833`. URL: `https://doi.org/10.1038/5833`.

[61] Tsaffrir Zor et al. “Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of c-Myb”. In: *Journal of Molecular Biology* 337.3 (2004), pp. 521–534. ISSN: 0022-2836. DOI: `https://doi.org/10.1016/j.jmb.2004.01.038`. URL: `https://www.sciencedirect.com/science/article/pii/S002228360400097X`.

[62] Jörg Gsponer et al. “Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation”. In: *Science* 322.5906 (2008), pp. 1365–1368. ISSN: 0036-8075. DOI: `10.1126/science.1163581`. URL: `https://science.sciencemag.org/content/322/5906/1365`.

[63] Mark O. Collins et al. "Phosphoproteomic Analysis of the Mouse Brain Cytosol Reveals a Predominance of Protein Phosphorylation in Regions of Intrinsic Sequence Disorder". In: *Molecular & Cellular Proteomics* 7.7 (2008), pp. 1331–1348. ISSN: 1535-9476. DOI: `https://doi.org/10.1074/mcp.M700564-MCP200`. URL: `https://www.sciencedirect.com/science/article/pii/S1535947620313359`.

[64] G. Manning et al. "The Protein Kinase Complement of the Human Genome". In: *Science* 298.5600 (2002), pp. 1912–1934. ISSN: 0036-8075. DOI: `10.1126/science.1075762`. URL: `https://science.sciencemag.org/content/298/5600/1912`.

[65] Alberto Paradela and Juan Pablo Albar. "Advances in the Analysis of Protein Phosphorylation". In: *Journal of Proteome Research* 7.5 (2008). PMID: 18327898, pp. 1809–1818. DOI: `10.1021/pr7006544`. URL: `https://doi.org/10.1021/pr7006544`.

[66] Tony Hunter. "Why nature chose phosphate to modify proteins". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1602 (2012), pp. 2513–2516. DOI: `10.1098/rstb.2012.0013`. URL: `https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2012.0013`.

[67] P J Kennelly and M Potts. "Fancy meeting you here! A fresh look at "prokaryotic" protein phosphorylation". In: *Journal of Bacteriology* 178.16 (1996), pp. 4759–4764. DOI: `10.1128/jb.178.16.4759-4764.1996`. URL: `https://journals.asm.org/doi/abs/10.1128/jb.178.16.4759-4764.1996`.

[68] Jun X Yan et al. "Protein phosphorylation: technologies for the identification of phosphoamino acids". In: *Journal of Chromatography A* 808.1 (1998), pp. 23–41. ISSN: 0021-9673. DOI: `https://doi.org/10.1016/S0021-9673(98)00115-0`. URL: `https://www.sciencedirect.com/science/article/pii/S0021967398001150`.

[69] Jung-Min Kee and Tom W. Muir. "Chasing Phosphohistidine, an Elusive Sibling in the Phosphoamino Acid Family". In: *ACS Chemical Biology* 7.1 (2012). PMID: 22148577, pp. 44–51. DOI: `10.1021/cb200445w`. URL: `https://doi.org/10.1021/cb200445w`.

[70] Michael Zachariou et al. "Potentiometric Investigations into the Acid-Base and Metal Ion Binding Properties of Immobilized Metal Ion Affinity Chromatographic (IMAC) Adsorbents". In: *The Journal of Physical Chemistry* 100.30 (1996), pp. 12680–12690. DOI: `10.1021/jp9601476`. URL: `https://doi.org/10.1021/jp9601476`.

[71] Maciej Śmiechowski. "Theoretical pKa prediction of O-phosphoserine in aqueous solution". In: *Chemical Physics Letters* 501.1 (2010), pp. 123–129. ISSN: 0009-2614. DOI: `https://doi.org/10.1016/j.cplett.2010.10.063`. URL: `https://www.sciencedirect.com/science/article/pii/S0009261410014442`.

[72] Daniel J. Mandell et al. "Strengths of Hydrogen Bonds Involving Phosphorylated Amino Acid Side Chains". In: *Journal of the American Chemical Society* 129.4 (2007). PMID: 17243818, pp. 820–827. DOI: `10.1021/ja063019w`. URL: `https://doi.org/10.1021/ja063019w`.

[73] Paulette Mhawech. "14-3-3 proteins—an update". In: *Cell Research* 15.4 (Apr. 2005), pp. 228–236. ISSN: 1748-7838. DOI: `10.1038/sj.cr.7290291`. URL: `https://doi.org/10.1038/sj.cr.7290291`.

[74] Alexandra K. Gardino, Stephen J. Smerdon, and Michael B. Yaffe. *Structural determinants of 14-3-3 binding specificities and regulation of subcellular localization of 14-3-3-ligand complexes: A comparison of the X-ray crystal structures of all human 14-3-3 isoforms*. June 2006. DOI: `10.1016/j.semcancer.2006.03.007`.

[75] Heiko Hermeking and Anne Benzinger. "14-3-3 proteins in cell cycle regulation". In: *Seminars in Cancer Biology* 16.3 (2006). 14-3-3 Proteins in Cancer, pp. 183–192. ISSN: 1044-579X. DOI: `https://doi.org/10.1016/j.semcancer.2006.03.002`. URL: `https://www.sciencedirect.com/science/article/pii/S1044579X06000265`.

[76] Erik Wilker and Michael B. Yaffe. "14-3-3 Proteins—a focus on cancer and human disease". In: *Journal of Molecular and Cellular Cardiology* 37.3 (2004), pp. 633–642. ISSN: 0022-2828. DOI: `https://doi.org/10.1016/j.yjmcc.2004.04.015`. URL: `https://www.sciencedirect.com/science/article/pii/S0022282804001178`.

[77] Guri Tzivion et al. "14-3-3 proteins as potential oncogenes". In: *Seminars in Cancer Biology* 16.3 (2006). 14-3-3 Proteins in Cancer, pp. 203–213. ISSN: 1044-579X. DOI: `https://doi.org/10.1016/j.semcancer.2006.03.004`. URL: `https://www.sciencedirect.com/science/article/pii/S1044579X06000289`.

[78] Christopher L. Neal and Dihua Yu. "14-3-3ζ as a prognostic marker and therapeutic target for cancer". In: *Expert Opinion on Therapeutic Targets* 14.12 (2010), pp. 1343–1354. DOI: `10.1517/14728222.2010.531011`. URL: `https://doi.org/10.1517/14728222.2010.531011`.

[79] Alyson K. Freeman and Deborah K. Morrison. "14-3-3 Proteins: Diverse functions in cell proliferation and cancer progression". In: *Seminars in Cell & Developmental Biology* 22.7 (2011), pp. 681–687. ISSN: 1084-9521. DOI: `https://doi.org/10.1016/j.semcdb.2011.08.009`. URL: `https://www.sciencedirect.com/science/article/pii/S1084952111001236`.

[80] Anne T. Ferguson et al. "High frequency of hypermethylation at the 14-3-3 σ locus leads to gene silencing in breast cancer". In: *Proceedings of the National Academy of Sciences* 97.11 (2000), pp. 6049–6054. ISSN: 0027-8424. DOI: `10.1073/pnas.100566997`. URL: `https://www.pnas.org/content/97/11/6049`.

[81] SeungSang Ko et al. "The Role and Regulatory Mechanism of 14-3-3 Sigma in Human Breast Cancer". In: *J Breast Cancer* 17.3 (Sept. 2014), pp. 207–218. ISSN: 1738-6756. DOI: `10.4048/jbc.2014.17.3.207`. URL: `https://doi.org/10.4048/jbc.2014.17.3.207`.

[82] Afshin Khorrami et al. "The functional significance of 14-3-3 proteins in cancer: focus on lung cancer". In: *Hormone Molecular Biology and Clinical Investigation* 32.3 (2017), p. 20170032. DOI: `10.1515/hmbci-2017-0032`. URL: `https://doi.org/10.1515/hmbci-2017-0032`.

[83] Jingyue Sun et al. "The survival analysis and oncogenic effects of CFP1 and 14-3-3 expression on gastric cancer". In: *Cancer Cell International* 19.1 (Aug. 2019), p. 225. ISSN: 1475-2867. DOI: `10.1186/s12935-019-0946-3`. URL: `https://doi.org/10.1186/s12935-019-0946-3`.

[84] Jiejie Qin et al. "Autoantibody against 14-3-3 zeta: a serological marker in detection of gastric cancer". In: *Journal of Cancer Research and Clinical Oncology* 145.5 (May 2019), pp. 1253–1262. ISSN: 1432-1335. DOI: `10.1007/s00432-019-02884-5`. URL: `https://doi.org/10.1007/s00432-019-02884-5`.

[85] Marta Falcicchio et al. "Regulation of p53 by the 14-3-3 protein interaction network: new opportunities for drug discovery in cancer". In: *Cell Death Discovery* 6.1 (Nov. 2020), p. 126. ISSN: 2058-7716. DOI: `10.1038/s41420-020-00362-3`. URL: `https://doi.org/10.1038/s41420-020-00362-3`.

[86] Daniela Berg, Carsten Holzmann, and Olaf Riess. "14-3-3 proteins in the nervous system". In: *Nature Reviews Neuroscience* 4.9 (Sept. 2003), pp. 752–762. ISSN: 1471-0048. DOI: `10.1038/nrn1197`. URL: `https://doi.org/10.1038/nrn1197`.

[87] Wei wen Ge et al. "14-3-3 protein binds to the low molecular weight neurofilament (NFL) mRNA 3′ UTR". In: *Molecular and Cellular Neuroscience* 34.1 (2007), pp. 80–87. ISSN: 1044-7431. DOI: `https://doi.org/10.1016/j.mcn.2006.10.001`. URL: `https://www.sciencedirect.com/science/article/pii/S1044743106002132`.

[88] Petra Steinacker, Alastair Aitken, and Markus Otto. "14-3-3 proteins in neurodegeneration". In: *Seminars in Cell & Developmental Biology* 22.7 (2011), pp. 696–704. ISSN: 1084-9521. DOI: `https://doi.org/10.1016/j.semcdb.2011.08.005`. URL: `https://www.sciencedirect.com/science/article/pii/S1084952111001194`.

[89] Xiaolan Fan et al. "14-3-3 Proteins Are on the Crossroads of Cancer, Aging, and Age-Related Neurodegenerative Disease". In: *International Journal of Molecular Sciences* 20.14 (2019). ISSN: 1422-0067. DOI: `10.3390/ijms20143518`. URL: `https://www.mdpi.com/1422-0067/20/14/3518`.

[90] Kavitha Ganesan Nathan and Sunil K. Lal. "The Multifarious Role of 14-3-3 Family of Proteins in Viral Replication". In: *Viruses* 12.4 (Apr. 2020), p. 436. ISSN: 1999-4915. DOI: `10.3390/v12040436`. URL: `https://doi.org/10.3390/v12040436`.

[91] Jiaqi Liu et al. "The role of 14-3-3 proteins in cell signalling pathways and virus infection". In: *Journal of Cellular and Molecular Medicine* 25.9 (2021), pp. 4173–4182. DOI: `https://doi.org/10.1111/jcmm.16490`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/jcmm.16490`.

[92] Hiroshi Aoki et al. "Hepatitis C Virus Core Protein Interacts with 14-3-3 Protein and Activates the Kinase Raf-1". In: *Journal of Virology* 74.4 (2000), pp. 1736–1741. DOI: `10.1128/JVI.74.4.1736-1741.2000`. URL: `https://journals.asm.org/doi/abs/10.1128/JVI.74.4.1736-1741.2000`.

[93] Milan Surjit et al. "The Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Protein Is Phosphorylated and Localizes in the Cytoplasm by 14-3-3-Mediated Translocation". In: *Journal of Virology* 79.17 (2005), pp. 11476–11486. DOI: `10.1128/JVI.79.17.11476-11486.2005`. URL: `https://journals.asm.org/doi/abs/10.1128/JVI.79.17.11476-11486.2005`.

[94] Kristina V. Tugaeva et al. "The Mechanism of SARS-CoV-2 Nucleocapsid Protein Recognition by the Human 14-3-3 Proteins". In: *Journal of Molecular Biology* 433.8 (2021), p. 166875. ISSN: 0022-2836. DOI: `https://doi.org/10.1016/j.jmb.2021.166875`. URL: `https://www.sciencedirect.com/science/article/pii/S0022283621000693`.

[95] H.Y. Lim Tung and Pierre Limtung. “Mutations in the phosphorylation sites of SARS-CoV-2 encoded nucleocapsid protein and structure model of sequestration by protein 14-3-3”. In: *Biochemical and Biophysical Research Communications* 532.1 (2020), pp. 134–138. ISSN: 0006-291X. DOI: `https://doi.org/10.1016/j.bbrc.2020.08.024`. URL: `https://www.sciencedirect.com/science/article/pii/S0006291X20315758`.

[96] Xiaowen Yang et al. “Structural basis for protein–protein interactions in the 14-3-3 protein family”. In: *Proceedings of the National Academy of Sciences* 103.46 (2006), pp. 17237–17242. ISSN: 0027-8424. DOI: `10.1073/pnas.0605779103`. URL: `https://www.pnas.org/content/103/46/17237`.

[97] Alastair Aitken. *14-3-3 and Its Possible Role in Co-Ordinating Multiple Signalling Pathways*. Sept. 1996. DOI: `10.1016/0962-8924(96)10029-5`.

[98] Haian Fu, Romesh R Subramanian, and Shane C Masters. “14-3-3 Proteins: Structure, Function, and Regulation”. In: *Annual Review of Pharmacology and Toxicology* 40.1 (2000), pp. 617–647. DOI: `10.1146/annurev.pharmtox.40.1.617`. URL: `https://doi.org/10.1146/annurev.pharmtox.40.1.617`.

[99] Joanna M. Woodcock et al. “Role of salt bridges in the dimer interface of 14-3-3ζ in dimer dynamics, N-terminal α-helical order, and molecular chaperone activity”. In: *Journal of Biological Chemistry* 293.1 (2018), pp. 89–99. ISSN: 1083351X. DOI: `10.1074/jbc.M117.801019`. URL: `http://www.jbc.org/`.

[100] Dave Bridges and Greg B. G. Moorhead. “14-3-3 Proteins: A Number of Functions for a Numbered Protein”. In: *Science Signaling* 2005.296 (2005), re10–re10. DOI: `10.1126/stke.2962005re10`. URL: `https://stke.sciencemag.org/content/2005/296/re10`.

[101] Lixin Zhang et al. “Raf-1 kinase and exoenzyme S interact with 14-3-3ζ through a common site involving lysine 49”. In: *Journal of Biological Chemistry* 272.21 (May 1997), pp. 13717–13724. ISSN: 00219258. DOI: `10.1074/jbc.272.21.13717`. URL: `http://www-jbc.stanford.edu/jbc/`.

[102] H. Wang et al. “Mutations in the hydrophobic surface of an amphiphatic groove of 14-3-3ζ disrupt its interaction with Raf-1 kinase”. In: *Journal of Biological Chemistry* 273.26 (June 1998), pp. 16297–16304. ISSN: 00219258. DOI: `10.1074/jbc.273.26.16297`. URL: `http://www.jbc.org`.

[103] C. Petosa et al. “14-3-3ζ Binds a Phosphorylated Raf Peptide and an Unphosphorylated Peptide Via Its Conserved Amphiphatic Groove”. In: *Journal of Biological Chemistry* 273.26 (1998), pp. 16305–16310. ISSN: 00219258. DOI: `10.1074/jbc.273.26.16305`.

[104] Michael B. Yaffe et al. “The structural basis for 14-3-3:phosphopeptide binding specificity”. In: *Cell* 91.7 (Dec. 1997), pp. 961–971. ISSN: 00928674. DOI: `10.1016/S0092-8674(00)80487-0`.

[105] Surajit Ganguly et al. “Melatonin synthesis: 14-3-3-Dependent activation and inhibition of arylalkylamine N-acetyltransferase mediated by phosphoserine-205”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.4 (Jan. 2005), pp. 1222–1227. ISSN: 00278424. DOI: `10.1073/pnas.0406871102`. URL: `/pmc/articles/PMC544185/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC544185/`.

[106] Kristina V. Tugaeva et al. "Concatenation of 14-3-3 with partner phosphoproteins as a tool to study their interaction". In: *Scientific Reports* 9.1 (2019). ISSN: 20452322. DOI: `10.1038/s41598-019-50941-3`. URL: `https://doi.org/10.1038/s41598-019-50941-3`.

[107] Alastair Aitken. "14-3-3 proteins: A historic overview". In: *Seminars in Cancer Biology* 16.3 (2006), pp. 162–172. ISSN: 1044579X. DOI: `10.1016/j.semcancer.2006.03.005`.

[108] Shogo Sato et al. "Metabolite Regulation of Nuclear Localization of Carbohydrate-response Element-binding Protein (ChREBP): ROLE OF AMP AS AN ALLOSTERIC INHIBITOR*". In: *Journal of Biological Chemistry* 291.20 (2016), pp. 10515–10527. ISSN: 0021-9258. DOI: `https://doi.org/10.1074/jbc.M115.708982`. URL: `https://www.sciencedirect.com/science/article/pii/S0021925820400870`.

[109] Eline Sijbesma et al. "Structure-based evolution of a promiscuous inhibitor to a selective stabilizer of protein-protein interactions". In: *Nature communications* 11.1 (Aug. 2020), pp. 3954–3954. ISSN: 2041-1723. DOI: `10.1038/s41467-020-17741-0`. URL: `https://doi.org/10.1038/s41467-020-17741-0`.

[110] Christian Ottmann et al. "Phosphorylation-independent interaction between 14-3-3 and exoenzyme S: From structure to pathogenesis". In: *EMBO Journal* 26.3 (Feb. 2007), pp. 902–913. ISSN: 02614189. DOI: `10.1038/sj.emboj.7601530`. URL: `http://emboj.embopress.org/cgi/doi/10.1038/sj.emboj.7601530`.

[111] Ruhangiz T Kilani et al. "Detection of high levels of 2 specific isoforms of 14-3-3 proteins in synovial fluid from patients with joint inflammation." In: *The Journal of Rheumatology* 34.8 (2007), pp. 1650–1657. ISSN: 0315-162X. URL: `https://www.jrheum.org/content/34/8/1650`.

[112] Walter P. Maksymowych et al. "14-3-3η is a novel mediator associated with the pathogenesis of rheumatoid arthritis and joint damage". In: *Arthritis Research & Therapy* 16.2 (Apr. 2014), R99. ISSN: 1478-6354. DOI: `10.1186/ar4547`. URL: `https://doi.org/10.1186/ar4547`.

[113] Tingting Zeng and Liming Tan. "14-3-3η protein: a promising biomarker for rheumatoid arthritis". In: *Biomarkers in Medicine* 12.8 (2018). PMID: 30022679, pp. 917–925. DOI: `10.2217/bmm-2017-0385`. URL: `https://doi.org/10.2217/bmm-2017-0385`.

[114] Tingting Zeng et al. "14-3-3η Protein in Rheumatoid Arthritis: Promising Diagnostic Marker and Independent Risk Factor for Osteoporosis". In: *Laboratory Medicine* 51.5 (Feb. 2020), pp. 529–539. ISSN: 0007-5027. DOI: `10.1093/labmed/lmaa001`. URL: `https://doi.org/10.1093/labmed/lmaa001`.

[115] Dina Anas Abdel Hai Hussin et al. "The significance of serum 14-3-3η level in rheumatoid arthritis patients". In: *Clinical Rheumatology* 40.6 (June 2021), pp. 2193–2202. ISSN: 1434-9949. DOI: `10.1007/s10067-020-05524-3`. URL: `https://doi.org/10.1007/s10067-020-05524-3`.

[116] C. G. Lee et al. "Roles of 14-3-3η in mitotic progression and its potential use as a therapeutic target for cancers". In: *Oncogene* 32.12 (Mar. 2013), pp. 1560–1569. ISSN: 1476-5594. DOI: `10.1038/onc.2012.170`. URL: `https://doi.org/10.1038/onc.2012.170`.

[117] Mary B. Kennedy. "Origin of PDZ (DHR, GLGF) domains". In: *Trends in Biochemical Sciences* 20.9 (1995), p. 350. ISSN: 0968-0004. DOI: `https://doi.org/10.1016/S0968-0004(00)89074-X`. URL: `https://www.sciencedirect.com/science/article/pii/S096800040089074X`.

[118] Declan A. Doyle et al. "Crystal Structures of a Complexed and Peptide-Free Membrane Protein–Binding Domain: Molecular Basis of Peptide Recognition by PDZ". In: *Cell* 85.7 (1996), pp. 1067–1076. ISSN: 0092-8674. DOI: `https://doi.org/10.1016/S0092-8674(00)81307-0`. URL: `https://www.sciencedirect.com/science/article/pii/S0092867400813070`.

[119] Christopher P. Ponting. "Evidence for PDZ domains in bacteria, yeast, and plants". In: *Protein Science* 6.2 (1997), pp. 464–468. DOI: `https://doi.org/10.1002/pro.5560060225`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.5560060225`.

[120] Albert Y. Hung and Morgan Sheng. "PDZ Domains: Structural Modules for Protein Complex Assembly". In: *Journal of Biological Chemistry* 277.8 (2002), pp. 5699–5702. ISSN: 0021-9258. DOI: `https://doi.org/10.1074/jbc.R100065200`. URL: `https://www.sciencedirect.com/science/article/pii/S0021925819823281`.

[121] Ho-Jin Lee and Jie J. Zheng. "PDZ domains and their binding partners: structure, specificity, and modification". In: *Cell Communication and Signaling* 8.1 (May 2010), p. 8. ISSN: 1478-811X. DOI: `10.1186/1478-811X-8-8`. URL: `https://doi.org/10.1186/1478-811X-8-8`.

[122] Craig C Garner, Joanne Nash, and Richard L Huganir. "PDZ domains in synapse assembly and signalling". In: *Trends in Cell Biology* 10.7 (2000), pp. 274–280. ISSN: 0962-8924. DOI: `https://doi.org/10.1016/S0962-8924(00)01783-9`. URL: `https://www.sciencedirect.com/science/article/pii/S0962892400017839`.

[123] Morgan Sheng and Carlo Sala. "PDZ Domains and the Organization of Supramolecular Complexes". In: *Annual Review of Neuroscience* 24.1 (2001). PMID: 11283303, pp. 1–29. DOI: `10.1146/annurev.neuro.24.1.1`. URL: `https://doi.org/10.1146/annurev.neuro.24.1.1`.

[124] Baruch Z. Harris and Wendell A. Lim. "Mechanism and role of PDZ domains in signaling complex assembly". In: *Journal of Cell Science* 114.18 (2001), pp. 3219–3231. ISSN: 0021-9533. URL: `https://jcs.biologists.org/content/114/18/3219`.

[125] Jing-Song Fan and Mingjie Zhang. "Signaling complex organization by PDZ domain proteins". In: *Neuro-Signals* 11.6 (2002), 315—321. ISSN: 1424-862X. DOI: `10.1159/000068256`. URL: `https://doi.org/10.1159/000068256`.

[126] Mark R. Spaller. "Act Globally, Think Locally: Systems Biology Addresses the PDZ Domain". In: *ACS Chemical Biology* 1.4 (2006). PMID: 17163673, pp. 207–210. DOI: `10.1021/cb600191y`. URL: `https://doi.org/10.1021/cb600191y`.

[127] Nikolaj R. Christensen et al. "PDZ Domains as Drug Targets". In: *Advanced Therapeutics* 2.7 (2019), p. 1800143. DOI: `https://doi.org/10.1002/adtp.201800143`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/adtp.201800143`.

[128] Jeanine F. Amacher et al. "Specificity in PDZ-peptide interaction networks: Computational analysis and review". In: *Journal of Structural Biology: X* 4 (2020), p. 100022. ISSN: 2590-1524. DOI: `https://doi.org/10.1016/j.yjsbx.2020.100022`. URL: `https://www.sciencedirect.com/science/article/pii/S2590152420300040`.

[129] M. Thomas et al. "Human papillomaviruses, cervical cancer and cell polarity". In: *Oncogene* 27.55 (Nov. 2008), pp. 7018–7030. ISSN: 1476-5594. DOI: `10.1038/onc.2008.351`. URL: `https://doi.org/10.1038/onc.2008.351`.

[130] David Pim et al. "Human papillomaviruses and the specificity of PDZ domain targeting". In: *The FEBS Journal* 279.19 (2012), pp. 3530–3537. DOI: `https://doi.org/10.1111/j.1742-4658.2012.08709.x`. URL: `https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1742-4658.2012.08709.x`.

[131] Claire D. James and Sally Roberts. "Viral Interactions with PDZ Domain-Containing Proteins—An Oncogenic Trait?" In: *Pathogens* 5.1 (2016). ISSN: 2076-0817. DOI: `10.3390/pathogens5010008`. URL: `https://www.mdpi.com/2076-0817/5/1/8`.

[132] Dante Barreda et al. "PDZ proteins are expressed and regulated in antigen-presenting cells and are targets of influenza A virus". In: *Journal of Leukocyte Biology* 103.4 (2018), pp. 731–738. DOI: `https://doi.org/10.1002/JLB.4AB0517-184R`. URL: `https://jlb.onlinelibrary.wiley.com/doi/abs/10.1002/JLB.4AB0517-184R`.

[133] Luis H. Gutiérrez-González and Teresa Santos-Mendoza. "Viral targeting of PDZ polarity proteins in the immune system as a potential evasion mechanism". In: *The FASEB Journal* 33.10 (2019), pp. 10607–10617. DOI: `https://doi.org/10.1096/fj.201900518R`. URL: `https://faseb.onlinelibrary.wiley.com/doi/abs/10.1096/fj.201900518R`.

[134] Mariano Genera et al. "Structural and functional characterization of the PDZ domain of the human phosphatase PTPN3 and its interaction with the human papillomavirus E6 oncoprotein". In: *Scientific Reports* 9.1 (May 2019), p. 7438. DOI: `10.1038/s41598-019-43932-x`. URL: `https://doi.org/10.1038/s41598-019-43932-x`.

[135] Angelo Toto et al. "Comparing the binding properties of peptides mimicking the Envelope protein of SARS-CoV and SARS-CoV-2 to the PDZ domain of the tight junction-associated PALS1 protein". In: *Protein Science* 29.10 (2020), pp. 2038–2042. DOI: `https://doi.org/10.1002/pro.3936`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.3936`.

[136] Carlos Casta~no-Rodriguez et al. "Role of Severe Acute Respiratory Syndrome Coronavirus Viroporins E, 3a, and 8a in Replication and Pathogenesis". In: *mBio* 9.3 (2018), e02325–17. DOI: `10.1128/mBio.02325-17`. URL: `https://journals.asm.org/doi/abs/10.1128/mBio.02325-17`.

[137] Ariel Shepley-McTaggart et al. "SARS-CoV-2 Envelope (E) protein interacts with PDZ-domain-2 of host tight junction protein ZO1". In: *PLOS ONE* 16.6 (June 2021), pp. 1–10. DOI: `10.1371/journal.pone.0251955`. URL: `https://doi.org/10.1371/journal.pone.0251955`.

[138] Célia Caillet-Saguy et al. "Host PDZ-containing proteins targeted by SARS-CoV-2". In: *The FEBS Journal* (2021). DOI: `https://doi.org/10.1111/febs.15881`. URL: `https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/febs.15881`.

[139] Matthew Gallon et al. “A unique PDZ domain and arrestin-like fold interaction reveals mechanistic details of endocytic recycling by SNX27-retromer”. In: *Proceedings of the National Academy of Sciences* 111.35 (2014), E3604–E3613. ISSN: 0027-8424. DOI: `10.1073/pnas.1410552111`. URL: `https://www.pnas.org/content/111/35/E3604`.

[140] Thomas Bourgeron. “From the genetic architecture to synaptic plasticity in autism spectrum disorder”. In: *Nature Reviews Neuroscience* 16.9 (Sept. 2015), pp. 551–563. ISSN: 1471-0048. DOI: `10.1038/nrn3992`. URL: `https://doi.org/10.1038/nrn3992`.

[141] Holly A. F. Stessman et al. “Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases”. In: *Nature Genetics* 49.4 (Apr. 2017), pp. 515–526. ISSN: 1546-1718. DOI: `10.1038/ng.3792`. URL: `https://doi.org/10.1038/ng.3792`.

[142] Edward Y. S. Lin et al. “Potent PDZ-Domain PICK1 Inhibitors that Modulate Amyloid Beta-Mediated Synaptic Dysfunction”. In: *Scientific Reports* 8.1 (Sept. 2018), p. 13438. ISSN: 2045-2322. DOI: `10.1038/s41598-018-31680-3`. URL: `https://doi.org/10.1038/s41598-018-31680-3`.

[143] Vanitha Krishna Subbaiah et al. “PDZ domains: the building blocks regulating tumorigenesis”. In: *Biochemical Journal* 439.2 (Sept. 2011), pp. 195–205. ISSN: 0264-6021. DOI: `10.1042/BJ20110903`. URL: `https://doi.org/10.1042/BJ20110903`.

[144] J. Vaquero et al. “Role of the PDZ-scaffold protein NHERF1/EBP50 in cancer biology: from signaling regulation to clinical relevance”. In: *Oncogene* 36.22 (June 2017), pp. 3067–3079. ISSN: 1476-5594. DOI: `10.1038/onc.2016.462`. URL: `https://doi.org/10.1038/onc.2016.462`.

[145] Yongsheng Yu et al. “A PDZ Protein MDA-9/Syntenin: As a Target for Cancer Therapy”. In: *Computational and Structural Biotechnology Journal* 17 (2019), pp. 136–141. ISSN: 2001-0370. DOI: `https://doi.org/10.1016/j.csbj.2019.01.002`. URL: `https://www.sciencedirect.com/science/article/pii/S200103701830271X`.

[146] R. Leblanc et al. “Pharmacological inhibition of syntenin PDZ2 domain impairs breast cancer cell activities and exosome loading with syndecan and EpCAM cargo”. In: *Journal of Extracellular Vesicles* 10.2 (2020), e12039. DOI: `https://doi.org/10.1002/jev2.12039`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jev2.12039`.

[147] Raffi Tonikian et al. “A Specificity Map for the PDZ Domain Family”. In: *PLOS Biology* 6.9 (Sept. 2008), pp. 1–17. DOI: `10.1371/journal.pbio.0060239`. URL: `https://doi.org/10.1371/journal.pbio.0060239`.

[148] Kousik Kundu and Rolf Backofen. “Cluster based prediction of PDZ-peptide interactions”. In: *BMC Genomics* 15.1 (Jan. 2014), S5. ISSN: 1471-2164. DOI: `10.1186/1471-2164-15-S1-S5`. URL: `https://doi.org/10.1186/1471-2164-15-S1-S5`.

[149] Kazuhiko Maekawa et al. “Molecular cloning of a novel protein-tyrosine phosphatase containing a membrane-binding domain and GLGF repeats”. In: *FEBS Letters* 337.2 (Jan. 1994), pp. 200–206. ISSN: 0014-5793. DOI: `https://doi.org/10.1016/0014-5793(94)80273-4`. URL: `https://www.sciencedirect.com/science/article/pii/0014579394802734`.

[150] D. Banville et al. "A novel protein-tyrosine phosphatase with homology to both the cytoskeletal proteins of the band 4.1 family and junction-associated guanylate kinases." In: *Journal of Biological Chemistry* 269.35 (Sept. 1994), pp. 22320–22327. ISSN: 0021-9258. DOI: `https://doi.org/10.1016/S0021-9258(17)31792-1`. URL: `https://www.sciencedirect.com/science/article/pii/S0021925817317921`.

[151] Soha Mcheik et al. "Dual Role of the PTPN13 Tyrosine Phosphatase in Cancer". In: *Biomolecules* 10.12 (2020). ISSN: 2218-273X. DOI: `10.3390/biom10121659`. URL: `https://www.mdpi.com/2218-273X/10/12/1659`.

[152] Gilles Freiss and Dany Chalbos. "PTPN13/PTPL1: An Important Regulator of Tumor Aggressiveness". In: *Anti-Cancer Agents in Medicinal Chemistry* 11.1 (2011), pp. 78–88. ISSN: 1871-5206/1875-5992. DOI: `10.2174/187152011794941262`. URL: `http://www.eurekaselect.com/node/73435/article`.

[153] Takaaki Sato et al. "FAP-1: A Protein Tyrosine Phosphatase That Associates with Fas". In: *Science* 268.5209 (1995), pp. 411–415. ISSN: 00368075, 10959203. URL: `http://www.jstor.org/stable/2886598`.

[154] Edwin Cuppen et al. "PDZ Motifs in PTP-BL and RIL Bind to Internal Protein Segments in the LIM Domain Protein RIL". In: *Molecular Biology of the Cell* 9.3 (1998). PMID: 9487134, pp. 671–683. DOI: `10.1091/mbc.9.3.671`. URL: `https://doi.org/10.1091/mbc.9.3.671`.

[155] Natalia S. Sotelo et al. "PTEN–PDZ domain interactions: Binding of PTEN to PDZ domains of PTPN13". In: *Methods* 77-78 (2015). PTEN Function Methods, pp. 147–156. ISSN: 1046-2023. DOI: `https://doi.org/10.1016/j.ymeth.2014.10.017`. URL: `https://www.sciencedirect.com/science/article/pii/S104620231400334X`.

[156] Angelo Toto et al. "Understanding the role of phosphorylation in the binding mechanism of a PDZ domain". In: *Protein Engineering, Design and Selection* 30.1 (Oct. 2017), pp. 1–5. ISSN: 17410134. DOI: `10.1093/protein/gzw055`. URL: `https://academic.oup.com/peds/article-lookup/doi/10.1093/protein/gzw055`.

[157] Irina Dobrosotskaya, Rodney K. Guy, and Guy L. James. "MAGI-1, a Membrane-associated Guanylate Kinase with a Unique Arrangement of Protein-Protein Interaction Domains *". In: *Journal of Biological Chemistry* 272.50 (Dec. 1997), pp. 31589–31597. ISSN: 0021-9258. DOI: `10.1074/jbc.272.50.31589`. URL: `https://doi.org/10.1074/jbc.272.50.31589`.

[158] Nobuyuki Ide et al. "Localization of membrane-associated guanylate kinase (MAGI)-1/BAI-associated protein (BAP) 1 at tight junctions of epithelial cells". In: *Oncogene* 18.54 (Dec. 1999), pp. 7810–7815. ISSN: 1476-5594. DOI: `10.1038/sj.onc.1203153`. URL: `https://doi.org/10.1038/sj.onc.1203153`.

[159] Kevin M. Patrie et al. "Interaction of Two Actin-binding Proteins, Synaptopodin and α-Actinin-4, with the Tight Junction Protein MAGI-1*". In: *Journal of Biological Chemistry* 277.33 (2002), pp. 30183–30190. ISSN: 0021-9258. DOI: `https://doi.org/10.1074/jbc.M203072200`. URL: `https://www.sciencedirect.com/science/article/pii/S0021925818756895`.

[160] Richard P. Laura et al. "MAGI-1: A Widely Expressed, Alternatively Spliced Tight Junction Protein". In: *Experimental Cell Research* 275.2 (2002), pp. 155–170. ISSN: 0014-4827. DOI: `https://doi.org/10.1006/excr.2002.5475`. URL: `https://www.sciencedirect.com/science/article/pii/S0014482702954753`.

[161] Larissa Kotelevets et al. “Implication of the MAGI-1b/PTEN signalosome in stabilization of adherens junctions and suppression of invasiveness”. In: *The FASEB Journal* 19.1 (2005), pp. 115–117. DOI: `https://doi.org/10.1096/fj.04-1942fje`. URL: `https://faseb.onlinelibrary.wiley.com/doi/abs/10.1096/fj.04-1942fje`.

[162] Bego~na Alday-Parejo et al. “MAGI1, a New Potential Tumor Suppressor Gene in Estrogen Receptor Positive Breast Cancer”. In: *Cancers* 12.1 (2020). ISSN: 2072-6694. DOI: `10.3390/cancers12010223`. URL: `https://www.mdpi.com/2072-6694/12/1/223`.

[163] Andrew Chatraryamontri et al. “The BioGRID interaction database: 2017 update”. In: *Nucleic Acids Research* 45.D1 (Dec. 2016), pp. D369–D379. ISSN: 0305-1048. DOI: `10.1093/nar/gkw1102`. URL: `https://dx.doi.org/10.1093/nar/gkw1102`.

[164] Alexander Roth et al. “The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible”. In: *Nucleic Acids Research* 45.D1 (Oct. 2016), pp. D362–D368. ISSN: 0305-1048. DOI: `10.1093/nar/gkw937`. URL: `https://dx.doi.org/10.1093/nar/gkw937`.

[165] Hing-C. Wong et al. “Direct Binding of the PDZ Domain of Dishevelled to a Conserved Internal Sequence in the C-Terminal Region of Frizzled”. In: *Molecular Cell* 12.5 (2003), pp. 1251–1260. ISSN: 1097-2765. DOI: `https://doi.org/10.1016/S1097-2765(03)00427-1`. URL: `https://www.sciencedirect.com/science/article/pii/S1097276503004271`.

[166] Alexandra Klaus and Walter Birchmeier. “Wnt signalling and its impact on development and cancer”. In: *Nature Reviews Cancer* 8.5 (May 2008), pp. 387–398. ISSN: 1474-1768. DOI: `10.1038/nrc2389`. URL: `https://doi.org/10.1038/nrc2389`.

[167] Paul Polakis. “Drugging Wnt signalling in cancer”. In: *The EMBO Journal* 31.12 (2012), pp. 2737–2746. DOI: `https://doi.org/10.1038/emboj.2012.126`. URL: `https://www.embopress.org/doi/abs/10.1038/emboj.2012.126`.

[168] “Wnt signaling in stem and cancer stem cells”. In: *Current Opinion in Cell Biology* 25.2 (2013). Cell regulation, pp. 254–264. ISSN: 0955-0674. DOI: `https://doi.org/10.1016/j.ceb.2013.01.004`. URL: `https://www.sciencedirect.com/science/article/pii/S0955067413000057`.

[169] N. Kamdem et al. “Small-molecule inhibitors of the PDZ domain of Dishevelled proteins interrupt Wnt signalling”. In: *Magnetic Resonance* 2.1 (2021), pp. 355–374. DOI: `10.5194/mr-2-355-2021`. URL: `https://mr.copernicus.org/articles/2/355/2021/`.

[170] Thuan D. Bui et al. “cDNA Cloning of a Humandishevelled DVL-3Gene, Mapping to 3q27, and Expression in Human Breast and Colon Carcinomas”. In: *Biochemical and Biophysical Research Communications* 239.2 (1997), pp. 510–516. ISSN: 0006-291X. DOI: `https://doi.org/10.1006/bbrc.1997.7500`. URL: `https://www.sciencedirect.com/science/article/pii/S0006291X97975008`.

[171] Ashani T. Weeraratna et al. “Wnt5a signaling directly affects cell motility and invasion of metastatic melanoma”. In: *Cancer Cell* 1.3 (2002), pp. 279–288. ISSN: 1535-6108. DOI: `https://doi.org/10.1016/S1535-6108(02)00045-4`. URL: `https://www.sciencedirect.com/science/article/pii/S1535610802000454`.

[172] Kazutsugu Uematsu et al. "Activation of the Wnt pathway in non small cell lung cancer: evidence of dishevelled overexpression". In: *Oncogene* 22.46 (Oct. 2003), pp. 7218–7221. ISSN: 1476-5594. DOI: `10.1038/sj.onc.1206817`. URL: `https://doi.org/10.1038/sj.onc.1206817`.

[173] Kazutsugu Uematsu et al. "Wnt Pathway Activation in Mesothelioma". In: *Cancer Research* 63.15 (2003), pp. 4547–4551. ISSN: 0008-5472. URL: `https://cancerres.aacrjournals.org/content/63/15/4547`.

[174] Kazunori Mizutani et al. "Upregulation and Overexpression of DVL1, the Human Counterpart of the Drosophila Dishevelled Gene, in Prostate Cancer". In: *Tumori Journal* 91.6 (2005). PMID: 16457155, pp. 546–551. DOI: `10.1177/030089160509100616`. URL: `https://doi.org/10.1177/030089160509100616`.

[175] Ho-Jin Lee et al. "Sulindac Inhibits Canonical Wnt Signaling by Blocking the PDZ Domain of the Protein Dishevelled". In: *Angewandte Chemie International Edition* 48.35 (2009), pp. 6448–6452. DOI: `https://doi.org/10.1002/anie.200902981`. URL: `https://www.onlinelibrary.wiley.com/doi/abs/10.1002/anie.200902981`.

[176] M. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford Graduate Texts. OUP Oxford, 2010. ISBN: 9780191523465. URL: `https://books.google.de/books?id=Lo3JqcOpgrcC`.

[177] J.W. Gibbs. *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundations of Thermodynamics*. Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics. C. Scribner's sons, 1902. URL: `https://books.google.de/books?id=IGMSAAAAIAAJ`.

[178] Ralph Baierlein. *Thermal Physics*. Cambridge University Press, 1999. DOI: `10.1017/CBO9780511840227`.

[179] Deutsche Akademie der Wissenschaften zu Berlin. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin.* Vol. Jan-Mai 1882. Berlin: Deutsche Akademie der Wissenschaften zu Berlin, 1882, p. 656. URL: `https://www.biodiversitylibrary.org/item/93362`.

[180] Herman J. C. Berendsen. *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*. Cambridge University Press, 2007. DOI: `10.1017/CBO9780511815348`.

[181] Jeffrey Comer et al. "The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask". In: *The Journal of Physical Chemistry B* 119.3 (2015). PMID: 25247823, pp. 1129–1151. DOI: `10.1021/jp506633n`. URL: `https://doi.org/10.1021/jp506633n`.

[182] C. Chipot and A. Pohorille. *Free Energy Calculations - Theory and Applications in Chemistry and Biology*. SPRINGER SERIES IN CHEMICAL PHYSICS. Springer-Verlag Berlin Heidelberg, 2007. ISBN: 978-3-540-38447-2. URL: `http://www.springer.com/de/book/9783540384472`.

[183] Cameron Abrams and Giovanni Bussi. "Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration". In: *Entropy* 16.1 (2014). DOI: `10.3390/e16010163`. URL: `https://doi.org/10.3390/e16010163`.

[184] C. Jarzynski. "Nonequilibrium Equality for Free Energy Differences". In: *Phys. Rev. Lett.* 78 (14 Apr. 1997), pp. 2690–2693. DOI: `10.1103/PhysRevLett.78.2690`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.78.2690`.

[185] Sanghyun Park and Klaus Schulten. "Calculating potentials of mean force from steered molecular dynamics simulations". In: *The Journal of Chemical Physics* 120.13 (2004), pp. 5946–5961. DOI: 10.1063/1.1651473. URL: https://doi.org/10.1063/1.1651473.

[186] Gavin E. Crooks. "Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems". In: *Journal of Statistical Physics* 90.5 (Mar. 1998), pp. 1481–1487. ISSN: 1572-9613. DOI: 10.1023/A:1023208217925. URL: https://doi.org/10.1023/A:1023208217925.

[187] M. Born and R. Oppenheimer. "Zur Quantentheorie der Molekeln". In: *Annalen der Physik* 389.20 (1927), pp. 457–484. DOI: https://doi.org/10.1002/andp.19273892002. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19273892002.

[188] Errol G. Lewars. *Computational Chemistry*. Springer International Publishing, 2016. DOI: 10.1007/978-3-319-30916-3. URL: https://doi.org/10.1007/978-3-319-30916-3.

[189] Valeri Poltev. "Molecular Mechanics: Principles, History, and Current Status". In: *Handbook of Computational Chemistry*. Ed. by Jerzy Leszczynski. Dordrecht: Springer Netherlands, 2016, pp. 1–48. ISBN: 978-94-007-6169-8. DOI: 10.1007/978-94-007-6169-8_9-2. URL: https://doi.org/10.1007/978-94-007-6169-8_9-2.

[190] A. D. MacKerell et al. "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins". In: *The Journal of Physical Chemistry B* 102.18 (1998), pp. 3586–3616. DOI: 10.1021/jp973084f. URL: https://doi.org/10.1021/jp973084f.

[191] Alexander D. Mackerell Jr., Michael Feig, and Charles L. Brooks III. "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations". In: *Journal of Computational Chemistry* 25.11 (2004), pp. 1400–1415. DOI: https://doi.org/10.1002/jcc.20065. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20065.

[192] Jing Huang et al. "CHARMM36m: An improved force field for folded and intrinsically disordered proteins". In: *Nature Methods* 14.1 (2016), pp. 71–73. ISSN: 15487105. DOI: 10.1038/nmeth.4067. URL: http://dx.doi.org/10.1038/nmeth.4067.

[193] Nicholas Metropolis et al. "Equation of state calculations by fast computing machines". In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092. ISSN: 00219606. DOI: 10.1063/1.1699114. URL: https://doi.org/10.1063/1.1730376.

[194] Mark E. Tuckerman and Glenn J. Martyna. "Understanding Modern Molecular Dynamics: Techniques and Applications". In: *The Journal of Physical Chemistry B* 104.2 (2000), pp. 159–178. DOI: 10.1021/jp992433y. URL: https://doi.org/10.1021/jp992433y.

[195] W. Kohn. "Nobel Lecture: Electronic structure of matter—wave functions and density functionals". In: *Rev. Mod. Phys.* 71 (5 Oct. 1999), pp. 1253–1266. DOI: 10.1103/RevModPhys.71.1253. URL: https://link.aps.org/doi/10.1103/RevModPhys.71.1253.

[196] Dominik Marx and Michele Parrinello. "Ab initio path integral molecular dynamics: Basic ideas". In: *The Journal of Chemical Physics* 104.11 (1996), pp. 4077–4082. DOI: `10.1063/1.471221`. URL: `https://doi.org/10.1063/1.471221`.

[197] Loup Verlet. "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules". In: *Phys. Rev.* 159 (1 July 1967), pp. 98–103. DOI: `10.1103/PhysRev.159.98`. URL: `https://link.aps.org/doi/10.1103/PhysRev.159.98`.

[198] William C. Swope et al. "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters". In: *The Journal of Chemical Physics* 76.1 (1982), pp. 637–649. DOI: `10.1063/1.442716`. URL: `https://doi.org/10.1063/1.442716`.

[199] R.W. Hockney, S.P. Goel, and J.W. Eastwood. "A 10000 particle molecular dynamics model with long range forces". In: *Chemical Physics Letters* 21.3 (1973), pp. 589–591. ISSN: 0009-2614. DOI: `https://doi.org/10.1016/0009-2614(73)80315-X`. URL: `https://www.sciencedirect.com/science/article/pii/000926147380315X`.

[200] Gabriel J. Rocklin et al. "Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects". In: *Journal of Chemical Physics* 139.18 (Nov. 2013), p. 184103. ISSN: 00219606. DOI: `10.1063/1.4826261`. URL: `http://www.ncbi.nlm.nih.gov/pubmed/24320250http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3838431`.

[201] G. Makov and M. C. Payne. "Periodic boundary conditions in ab initio calculations". In: *Phys. Rev. B* 51 (7 Feb. 1995), pp. 4014–4022. DOI: `10.1103/PhysRevB.51.4014`. URL: `https://link.aps.org/doi/10.1103/PhysRevB.51.4014`.

[202] Marcus B. Kubitzki, Bert L. de Groot, and Daniel Seeliger. "Protein DynamicsProtein Dynamics: From Structure to Function". In: *From Protein Structure to Function with Bioinformatics*. Ed. by Daniel J. Rigden. Dordrecht: Springer Netherlands, 2017, pp. 393–425. ISBN: 978-94-024-1069-3. DOI: `10.1007/978-94-024-1069-3_12`. URL: `https://doi.org/10.1007/978-94-024-1069-3_12`.

[203] Regula Walser, Philippe H. Hünenberger, and Wilfred F. van Gunsteren. "Comparison of different schemes to treat long-range electrostatic interactions in molecular dynamics simulations of a protein crystal". In: *Proteins: Structure, Function, and Bioinformatics* 43.4 (2001), pp. 509–519. DOI: `https://doi.org/10.1002/prot.1062`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.1062`.

[204] Brock A. Luty, Ilario G. Tironi, and Wilfred F. van Gunsteren. "Lattice-sum methods for calculating electrostatic interactions in molecular simulations". In: *The Journal of Chemical Physics* 103.8 (1995), pp. 3014–3021. DOI: `10.1063/1.470490`. URL: `https://doi.org/10.1063/1.470490`.

[205] P. P. Ewald. "Die Berechnung optischer und elektrostatischer Gitterpotentiale". In: *Annalen der Physik* 369.3 (1921), pp. 253–287. DOI: `10.1002/andp.19213690304`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.19213690304`.

[206] R. W. Hockney and J. W. Eastwood. *Computer Simulation Using Particles*. 1981.

[207] S. W. de Leeuw et al. "Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants". In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 373.1752 (1980), pp. 27–56. DOI: `10.1098/rspa.1980.0135`. URL: `https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1980.0135`.

[208] Philippe H. Hünenberger. "Lattice-sum methods for computing electrostatic interactions in molecular simulations". In: *AIP Conference Proceedings* 492.1 (1999), pp. 17–83. DOI: `10.1063/1.1301521`. URL: `https://aip.scitation.org/doi/abs/10.1063/1.1301521`.

[209] Brock A. Luty et al. "A Comparison of Particle-Particle, Particle-Mesh and Ewald Methods for Calculating Electrostatic Interactions in Periodic Molecular Systems". In: *Molecular Simulation* 14.1 (1994), pp. 11–20. DOI: `10.1080/08927029408022004`. URL: `https://doi.org/10.1080/08927029408022004`.

[210] E.L. Pollock and Jim Glosli. "Comments on P3M, FMM, and the Ewald method for large periodic Coulombic systems". In: *Computer Physics Communications* 95.2 (1996), pp. 93 –110. ISSN: 0010-4655. DOI: `https://doi.org/10.1016/0010-4655(96)00043-4`. URL: `http://www.sciencedirect.com/science/article/pii/0010465596000434`.

[211] Markus Deserno and Christian Holm. "How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines". In: *The Journal of Chemical Physics* 109.18 (1998), pp. 7678–7693. DOI: `10.1063/1.477414`. URL: `https://doi.org/10.1063/1.477414`.

[212] Markus Deserno and Christian Holm. "How to mesh up Ewald sums. II. An accurate error estimate for the particle-particle-particle-mesh algorithm". In: *The Journal of Chemical Physics* 109.18 (1998), pp. 7694–7701. DOI: `10.1063/1.477415`. URL: `https://doi.org/10.1063/1.477415`.

[213] Mika A. Kastenholz and Philippe H. Hünenberger. "Influence of Artificial Periodicity and Ionic Strength in Molecular Dynamics Simulations of Charged Biomolecules Employing Lattice-Sum Methods". In: *The Journal of Physical Chemistry B* 108.2 (2004), pp. 774–788. DOI: `10.1021/jp0350924`. URL: `https://doi.org/10.1021/jp0350924`.

[214] Philippe H. Hünenberger and J. Andrew McCammon. "Ewald artifacts in computer simulations of ionic solvation and ion-ion interaction: A continuum electrostatics study". In: *Journal of Chemical Physics* 110.4 (1999), pp. 1856–1872. ISSN: 00219606. DOI: `10.1063/1.477873`. URL: `https://doi.org/10.1063/1.477873https://doi.org/10.1063/1.4826261https://doi.org/10.1063/1.2201698`.

[215] R. Eisenschitz and F. London. "Über das Verhältnis der van der Waalsschen Kräfte zu den homöopolaren Bindungskräften". In: *Zeitschrift für Physik* 60.7 (July 1930), pp. 491–527. ISSN: 0044-3328. DOI: `10.1007/BF01341258`. URL: `https://doi.org/10.1007/BF01341258`.

[216] F. London. "Zur Theorie und Systematik der Molekularkräfte". In: *Zeitschrift für Physik* 63.3 (Mar. 1930), pp. 245–279. ISSN: 0044-3328. DOI: `10.1007/BF01421741`. URL: `https://doi.org/10.1007/BF01421741`.

[217] J. E. Jones and Sydney Chapman. "On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 106.738 (1924), pp. 441–462. DOI: `10.1098/rspa.1924.0081`. URL: `https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1924.0081`.

[218] J. E. Jones and Sydney Chapman. "On the determination of molecular fields. —II. From the equation of state of a gas". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 106.738 (1924), pp. 463–477. DOI: `10.1098/rspa.1924.0082`. URL: `https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1924.0082`.

[219] J E Lennard-Jones. "Cohesion". In: *Proceedings of the Physical Society* 43.5 (Sept. 1931), pp. 461–482. DOI: `10.1088/0959-5309/43/5/301`. URL: `https://doi.org/10.1088/0959-5309/43/5/301`.

[220] Alexey Onufriev. "Chapter 7 - Implicit Solvent Models in Molecular Dynamics Simulations: A Brief Overview". In: ed. by Ralph A. Wheeler and David C. Spellmeyer. Vol. 4. Annual Reports in Computational Chemistry. Elsevier, 2008, pp. 125–137. DOI: `https://doi.org/10.1016/S1574-1400(08)00007-8`. URL: `https://www.sciencedirect.com/science/article/pii/S1574140008000078`.

[221] Martin Lísal, Jiří Kolafa, and Ivo Nezbeda. "An examination of the five-site potential (TIP5P) for water". In: *The Journal of Chemical Physics* 117.19 (2002), pp. 8892–8897. DOI: `10.1063/1.1514572`. URL: `https://doi.org/10.1063/1.1514572`.

[222] Hans W. Horn et al. "Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew". In: *The Journal of Chemical Physics* 120.20 (2004), pp. 9665–9678. DOI: `10.1063/1.1683075`. URL: `https://doi.org/10.1063/1.1683075`.

[223] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. "The missing term in effective pair potentials". In: *The Journal of Physical Chemistry* 91.24 (1987), pp. 6269–6271. DOI: `10.1021/j100308a038`. URL: `https://doi.org/10.1021/j100308a038`.

[224] William L. Jorgensen et al. "Comparison of simple potential functions for simulating liquid water". In: *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935. DOI: `10.1063/1.445869`. URL: `https://doi.org/10.1063/1.445869`.

[225] Saeed Izadi, Ramu Anandakrishnan, and Alexey V. Onufriev. "Building Water Models: A Different Approach". In: *The Journal of Physical Chemistry Letters* 5.21 (2014). PMID: 25400877, pp. 3863–3871. DOI: `10.1021/jz501780a`. URL: `https://doi.org/10.1021/jz501780a`.

[226] Michael W. Mahoney and William L. Jorgensen. "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions". In: *The Journal of Chemical Physics* 112.20 (2000), pp. 8910–8922. DOI: `10.1063/1.481505`. URL: `https://doi.org/10.1063/1.481505`.

[227] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J.C Berendsen. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". In: *Journal of Computational Physics* 23.3 (1977), pp. 327–341. ISSN: 0021-9991. DOI: `https://doi.org/10.1016/0021-9991(77)90098-5`. URL: `https://www.sciencedirect.com/science/article/pii/0021999177900985`.

[228] Hans C Andersen. "Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations". In: *Journal of Computational Physics* 52.1 (1983), pp. 24–34. ISSN: 0021-9991. DOI: `https://doi.org/10.1016/0021-9991(83)90014-1`. URL: `https://www.sciencedirect.com/science/article/pii/0021999183900141`.

[229] Shuichi Miyamoto and Peter A. Kollman. "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models". In: *Journal of Computational Chemistry* 13.8 (1992), pp. 952–962. DOI: `https://doi.org/10.1002/jcc.540130805`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540130805`.

[230] Dietmar Paschek, Ryan Day, and Angel E. García. "Influence of water–protein hydrogen bonding on the stability of Trp-cage miniprotein. A comparison between the TIP3P and TIP4P-Ew water models". In: *Phys. Chem. Chem. Phys.* 13 (44 2011), pp. 19840–19847. DOI: `10.1039/C1CP22110H`. URL: `http://dx.doi.org/10.1039/C1CP22110H`.

[231] Sander Boonstra, Patrick R. Onck, and Erik van der Giessen. "CHARMM TIP3P Water Model Suppresses Peptide Folding by Solvating the Unfolded State". In: *The Journal of Physical Chemistry B* 120.15 (2016). PMID: 27031562, pp. 3692–3698. DOI: `10.1021/acs.jpcb.6b01316`. URL: `https://doi.org/10.1021/acs.jpcb.6b01316`.

[232] Elisa Fadda and Robert J. Woods. "On the Role of Water Models in Quantifying the Binding Free Energy of Highly Conserved Water Molecules in Proteins: The Case of Concanavalin A". In: *Journal of Chemical Theory and Computation* 7.10 (2011). PMID: 26598169, pp. 3391–3398. DOI: `10.1021/ct200404z`. URL: `https://doi.org/10.1021/ct200404z`.

[233] Saeed Izadi, Boris Aguilar, and Alexey V. Onufriev. "Protein–Ligand Electrostatic Binding Free Energies from Explicit and Implicit Solvation". In: *Journal of Chemical Theory and Computation* 11.9 (2015). PMID: 26575935, pp. 4450–4459. DOI: `10.1021/acs.jctc.5b00483`. URL: `https://doi.org/10.1021/acs.jctc.5b00483`.

[234] Philippe H. Hünenberger. "Thermostat Algorithms for Molecular Dynamics Simulations". In: *Advanced Computer Simulation: Approaches for Soft Matter Sciences I*. Ed. by Christian Dr. Holm and Kurt Prof. Dr. Kremer. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 105–149. ISBN: 978-3-540-31558-2. DOI: `10.1007/b99427`. URL: `https://doi.org/10.1007/b99427`.

[235] Hans C. Andersen. "Molecular dynamics simulations at constant pressure and/or temperature". In: *The Journal of Chemical Physics* 72.4 (1980), pp. 2384–2393. DOI: `10.1063/1.439486`. URL: `https://doi.org/10.1063/1.439486`.

[236] Shūichi Nosé. "A molecular dynamics method for simulations in the canonical ensemble". In: *Molecular Physics* 52.2 (1984), pp. 255–268. DOI: `10.1080/00268978400101201`. URL: `https://doi.org/10.1080/00268978400101201`.

[237] William G. Hoover. "Canonical dynamics: Equilibrium phase-space distributions". In: *Phys. Rev. A* 31 (3 Mar. 1985), pp. 1695–1697. DOI: 10.1103/PhysRevA.31.1695. URL: https://link.aps.org/doi/10.1103/PhysRevA.31.1695.

[238] H. J. C. Berendsen et al. "Molecular dynamics with coupling to an external bath". In: *The Journal of Chemical Physics* 81.8 (1984), pp. 3684–3690. DOI: 10.1063/1.448118. URL: https://doi.org/10.1063/1.448118.

[239] Giovanni Bussi, Davide Donadio, and Michele Parrinello. "Canonical sampling through velocity rescaling". In: *The Journal of Chemical Physics* 126.1 (2007), p. 014101. DOI: 10.1063/1.2408420. URL: https://doi.org/10.1063/1.2408420.

[240] Giovanni Bussi and Michele Parrinello. "Stochastic thermostats: comparison of local and global schemes". In: *Computer Physics Communications* 179.1 (2008). Special issue based on the Conference on Computational Physics 2007, pp. 26–29. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2008.01.006. URL: https://www.sciencedirect.com/science/article/pii/S0010465508000106.

[241] Stephen C. Harvey, Robert K.-Z. Tan, and Thomas E. Cheatham III. "The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition". In: *Journal of Computational Chemistry* 19.7 (1998), pp. 726–740. DOI: https://doi.org/10.1002/(SICI)1096-987X(199805)19:7<726::AID-JCC4>3.0.CO;2-S. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-987X%28199805%2919%3A7%3C726%3A%3AAID-JCC4%3E3.0.CO%3B2-S.

[242] Jirasak Wong-ekkabut and Mikko Karttunen. "The good, the bad and the user in soft matter simulations". In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1858.10 (2016). Biosimulations of lipid membranes coupled to experiments, pp. 2529–2538. ISSN: 0005-2736. DOI: https://doi.org/10.1016/j.bbamem.2016.02.004. URL: https://www.sciencedirect.com/science/article/pii/S0005273616300347.

[243] Allison E. Howard and Peter A. Kollman. "Molecular dynamics studies of a DNA-binding protein: 1. A comparison of the trp repressor and trp aporepressor aqueous simulations". In: *Protein Science* 1.9 (1992), pp. 1173–1184. DOI: https://doi.org/10.1002/pro.5560010911. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.5560010911.

[244] Jeanmarie Guenot and Peter A. Kollman. "Conformational and energetic effects of truncating nonbonded interactions in an aqueous protein dynamics simulation". In: *Journal of Computational Chemistry* 14.3 (1993), pp. 295–311. DOI: https://doi.org/10.1002/jcc.540140306. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540140306.

[245] Koji Oda, Hiroh Miyagawa, and Kunihiro Kitamura. "How does the Electrostatic Force Cut-Off Generate Non-uniform Temperature Distributions in Proteins?" In: *Molecular Simulation* 16.1-3 (1996), pp. 167–177. DOI: 10.1080/08927029608024070. URL: https://doi.org/10.1080/08927029608024070.

[246] M. Lingenheil et al. "The "Hot-Solvent/Cold-Solute" Problem Revisited". In: *Journal of Chemical Theory and Computation* 4.8 (2008). PMID: 26631705, pp. 1293–1306. DOI: 10.1021/ct8000365. URL: https://doi.org/10.1021/ct8000365.

[247] Efrem Braun et al. "Best Practices for Foundations in Molecular Simulations [Article v1.0]". In: *Living Journal of Computational Molecular Science* 1.1 (Nov. 2018), p. 5957. DOI: `10.33011/livecoms.1.1.5957`.

[248] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein. "Constant pressure molecular dynamics algorithms". In: *The Journal of Chemical Physics* 101.5 (1994), pp. 4177–4189. DOI: `10.1063/1.467468`. URL: `https://doi.org/10.1063/1.467468`.

[249] Glenn J. Martyna et al. "Explicit reversible integrators for extended systems dynamics". In: *Molecular Physics* 87.5 (1996), pp. 1117–1157. DOI: `10.1080/00268979600100761`. URL: `https://doi.org/10.1080/00268979600100761`.

[250] M. Parrinello and A. Rahman. "Polymorphic transitions in single crystals: A new molecular dynamics method". In: *Journal of Applied Physics* 52.12 (1981), pp. 7182–7190. DOI: `10.1063/1.328693`. URL: `https://doi.org/10.1063/1.328693`.

[251] David E. Shaw et al. "Anton, a Special-Purpose Machine for Molecular Dynamics Simulation". In: *SIGARCH Comput. Archit. News* 35.2 (June 2007), 1–12. ISSN: 0163-5964. DOI: `10.1145/1273440.1250664`. URL: `https://doi.org/10.1145/1273440.1250664`.

[252] D. E. Shaw et al. "Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer". In: *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. 2014, pp. 41–53. DOI: `10.1109/SC.2014.9`.

[253] Albert C. Pan et al. "Atomic-level characterization of protein–protein association". In: *Proceedings of the National Academy of Sciences* 116.10 (2019), pp. 4244–4249. ISSN: 0027-8424. DOI: `10.1073/pnas.1815431116`. URL: `https://www.pnas.org/content/116/10/4244`.

[254] Gavin E. Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences". In: *Phys. Rev. E* 60 (3 Sept. 1999), pp. 2721–2726. DOI: `10.1103/PhysRevE.60.2721`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.60.2721`.

[255] E.A. Carter et al. "Constrained reaction coordinate dynamics for the simulation of rare events". In: *Chemical Physics Letters* 156.5 (1989), pp. 472 – 477. ISSN: 0009-2614. DOI: `https://doi.org/10.1016/S0009-2614(89)87314-2`. URL: `http://www.sciencedirect.com/science/article/pii/S0009261489873142`.

[256] Gerhard Hummer, Lawrence R. Pratt, and Angel E. García. "Free Energy of Ionic Hydration". In: *The Journal of Physical Chemistry* 100.4 (1996), pp. 1206–1215. DOI: `10.1021/jp951011v`. URL: `https://doi.org/10.1021/jp951011v`.

[257] B. R. A. Nijboer and Th W. Ruijgrok. "On the energy per particle in three- and two-dimensional Wigner lattices". In: *Journal of Statistical Physics* 53.1 (1988), pp. 361–382. ISSN: 1572-9613. DOI: `10.1007/BF01011562`. URL: `https://doi.org/10.1007/BF01011562`.

[258] B. Cichocki, B. U. Felderhof, and K. Hinsen. "Electrostatic interactions in periodic Coulomb and dipolar systems". In: *Phys. Rev. A* 39 (10 May 1989), pp. 5350–5358. DOI: `10.1103/PhysRevA.39.5350`. URL: `https://link.aps.org/doi/10.1103/PhysRevA.39.5350`.

[259] Gerhard Hummer and Dikeos Mario Soumpasis. "Correlations and free energies in restricted primitive model descriptions of electrolytes". In: *The Journal of Chemical Physics* 98.1 (1993), pp. 581–591. DOI: `10.1063/1.464600`. URL: `https://doi.org/10.1063/1.464600`.

[260] Yury N. Vorobjev and Jan Hermans. "A Critical Analysis of Methods of Calculation of a Potential in Simulated Polar Liquids: Strong Arguments in Favor of "Molecule-Based" Summation and of Vacuum Boundary Conditions in Ewald Summation". In: *The Journal of Physical Chemistry B* 103.46 (1999), pp. 10234–10242. DOI: `10.1021/jp984211q`. URL: `https://doi.org/10.1021/jp984211q`.

[261] Gerhard Hummer, Lawrence R. Pratt, and Angel E. García. "Molecular Theories and Simulation of Ions and Polar Molecules in Water". In: *The Journal of Physical Chemistry A* 102.41 (1998), pp. 7885–7895. DOI: `10.1021/jp982195r`. URL: `https://doi.org/10.1021/jp982195r`.

[262] Henry S. Ashbaugh and Robert H. Wood. "Effects of long-range electrostatic potential truncation on the free energy of ionic hydration". In: *The Journal of Chemical Physics* 106.19 (1997), pp. 8135–8139. DOI: `10.1063/1.473800`. URL: `https://doi.org/10.1063/1.473800`.

[263] Gerhard Hummer et al. "Electrostatic Potentials and Free Energies of Solvation of Polar and Charged Molecules". In: *The Journal of Physical Chemistry B* 101.16 (1997), pp. 3017–3020. DOI: `10.1021/jp964037a`. URL: `https://doi.org/10.1021/jp964037a`.

[264] Johan Åqvist and Tomas Hansson. "Analysis of Electrostatic Potential Truncation Schemes in Simulations of Polar Solvents". In: *The Journal of Physical Chemistry B* 102.19 (1998), pp. 3837–3840. DOI: `10.1021/jp973207w`. URL: `https://doi.org/10.1021/jp973207w`.

[265] Henry S. Ashbaugh, Shinichi Sakane, and Robert H. Wood. "Reply to Comment on "Electrostatic Potentials and Free Energies of Solvation of Polar and Charged Molecules"". In: *The Journal of Physical Chemistry B* 102.19 (1998), pp. 3844–3845. DOI: `10.1021/jp9804852`. URL: `https://doi.org/10.1021/jp9804852`.

[266] C. Satheesan Babu, Pei-Kun Yang, and Carmay Lim. "On the Charge and Molecule Based Summations of Solvent Electrostatic Potentials and the Validity of Electrostatic Linear Response in Water". In: *Journal of Biological Physics* 28.2 (2002), pp. 95–113. ISSN: 1573-0689. DOI: `10.1023/A:1019978119707`. URL: `https://doi.org/10.1023/A:1019978119707`.

[267] Vytautas Gapsys et al. "pmx: Automated protein structure and topology generation for alchemical perturbations". In: *Journal of Computational Chemistry* 36.5 (Feb. 2015), pp. 348–354. ISSN: 1096987X. DOI: `10.1002/jcc.23804`. URL: `http://doi.wiley.com/10.1002/jcc.23804`.

[268] Vytautas Gapsys et al. "Calculation of binding free energies". In: *Molecular Modeling of Proteins: Second Edition* 1215 (2014), pp. 173–209. DOI: `10.1007/978-1-4939-1465-4_9`. URL: `https://www.mpibpc.mpg.de/14921456/Gapsys{\_}2015{\_}Springer.pdf`.

[269] M.K. Gilson et al. “The statistical-thermodynamic basis for computation of binding affinities: a critical review”. In: *Biophysical Journal* 72.3 (1997), pp. 1047 –1069. ISSN: 0006-3495. DOI: `https://doi.org/10.1016/S0006-3495(97)78756-3`. URL: `http://www.sciencedirect.com/science/article/pii/S0006349597787563`.

[270] Stefan Boresch et al. “Absolute Binding Free Energies: A Quantitative Approach for Their Calculation”. In: *The Journal of Physical Chemistry B* 107.35 (2003), pp. 9535–9551. DOI: `10.1021/jp0217839`. URL: `https://doi.org/10.1021/jp0217839`.

[271] Ignacio J. General. “A note on the standard state’s binding free energy”. In: *Journal of Chemical Theory and Computation* 6.8 (2010), pp. 2520–2524. DOI: `10.1021/ct100255z`. URL: `https://pubs.acs.org/sharingguidelines`.

[272] Yuqing Deng and Benoît Roux. “Computations of standard binding free energies with molecular dynamics simulations”. In: *Journal of Physical Chemistry B* 113.8 (2009), pp. 2234–2246. ISSN: 15206106. DOI: `10.1021/jp807701h`.

[273] Johan Åqvist, Carmen Medina, and Jan-Erik Samuelsson. “A new method for predicting binding affinity in computer-aided drug design”. In: *Protein Engineering, Design and Selection* 7.3 (Mar. 1994), pp. 385–391. ISSN: 1741-0126. DOI: `10.1093/protein/7.3.385`. URL: `https://doi.org/10.1093/protein/7.3.385`.

[274] Tomas Hansson, John Marelius, and Johan Åqvist. “Ligand binding affinity prediction by linear interaction energy methods”. In: *Journal of Computer-Aided Molecular Design* 12.1 (Jan. 1998), pp. 27–35. ISSN: 1573-4951. DOI: `10.1023/A:1007930623000`. URL: `https://doi.org/10.1023/A:1007930623000`.

[275] Jayashree Srinivasan et al. “Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices”. In: *Journal of the American Chemical Society* 120.37 (1998), pp. 9401–9409. DOI: `10.1021/ja981844+`. URL: `https://doi.org/10.1021/ja981844+`.

[276] Peter A. Kollman et al. “Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models”. In: *Accounts of Chemical Research* 33.12 (2000). PMID: 11123888, pp. 889–897. DOI: `10.1021/ar000033j`. URL: `https://doi.org/10.1021/ar000033j`.

[277] Samuel Genheden and Ulf Ryde. “The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities”. In: *Expert opinion on drug discovery* 10.5 (June 2015), 449—461. ISSN: 1746-0441. DOI: `10.1517/17460441.2015.1032936`. URL: `https://europepmc.org/articles/PMC4487606`.

[278] Holger Gohlke and David A. Case. “Converging free energy estimates: MM-PB(GB)SA studies on the protein–protein complex Ras-Raf”. In: *Journal of Computational Chemistry* 25.2 (2004), pp. 238–250. DOI: `https://doi.org/10.1002/jcc.10379`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.10379`.

[279] Jessica M.J. Swanson, Richard H. Henchman, and J. Andrew McCammon. “Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy”. In: *Biophysical Journal* 86.1 (2004), pp. 67 –74. ISSN: 0006-3495. DOI: `https://doi.org/10.1016/S0006-3495(04)74084-9`. URL: `http://www.sciencedirect.com/science/article/pii/S0006349504740849`.

[280] Samuel Genheden and Ulf Ryde. “A comparison of different initialization protocols to obtain statistically independent molecular dynamics simulations”. In: *Journal of Computational Chemistry* 32.2 (2011), pp. 187–195. DOI: `https://doi.org/10.1002/jcc.21546`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.21546`.

[281] Huiyong Sun et al. “Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set”. In: *Phys. Chem. Chem. Phys.* 16 (31 2014), pp. 16719–16729. DOI: `10.1039/C4CP01388C`. URL: `http://dx.doi.org/10.1039/C4CP01388C`.

[282] Rafael C. Bernardi, Marcelo C.R. Melo, and Klaus Schulten. “Enhanced sampling techniques in molecular dynamics simulations of biological systems”. In: *Biochimica et Biophysica Acta (BBA) - General Subjects* 1850.5 (2015). Recent developments of molecular dynamics, pp. 872 –877. ISSN: 0304-4165. DOI: `https://doi.org/10.1016/j.bbagen.2014.10.019`. URL: `http://www.sciencedirect.com/science/article/pii/S0304416514003559`.

[283] Cameron Abrams and Giovanni Bussi. “Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration”. In: *Entropy* 16.1 (2014), pp. 163–199. ISSN: 1099-4300. DOI: `10.3390/e16010163`. URL: `https://www.mdpi.com/1099-4300/16/1/163`.

[284] Manuel Luitz et al. “Exploring biomolecular dynamics and interactions using advanced sampling methods”. In: *Journal of Physics: Condensed Matter* 27.32 (July 2015), p. 323101. DOI: `10.1088/0953-8984/27/32/323101`. URL: `https://doi.org/10.1088/0953-8984/27/32/323101`.

[285] Till Siebenmorgen and Martin Zacharias. “Computational prediction of protein–protein binding affinities”. In: *WIREs Computational Molecular Science* 10.3 (2020), e1448. DOI: `https://doi.org/10.1002/wcms.1448`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1448`.

[286] Johannes Kästner. “Umbrella sampling”. In: *WIREs Computational Molecular Science* 1.6 (2011), pp. 932–942. DOI: `https://doi.org/10.1002/wcms.66`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.66`.

[287] Jan Hermans. “Simple analysis of noise and hysteresis in (slow-growth) free energy simulations”. In: *The Journal of Physical Chemistry* 95.23 (1991), pp. 9029–9032. DOI: `10.1021/j100176a002`. URL: `https://doi.org/10.1021/j100176a002`.

[288] G.M. Torrie and J.P. Valleau. “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. In: *Journal of Computational Physics* 23.2 (1977), pp. 187 –199. ISSN: 0021-9991. DOI: `https://doi.org/10.1016/0021-9991(77)90121-8`. URL: `http://www.sciencedirect.com/science/article/pii/0021999177901218`.

[289] Mihaly Mezei. “Adaptive umbrella sampling: Self-consistent determination of the non-Boltzmann bias”. In: *Journal of Computational Physics* 68.1 (1987), pp. 237 –248. ISSN: 0021-9991. DOI: `https://doi.org/10.1016/0021-9991(87)90054-4`. URL: `http://www.sciencedirect.com/science/article/pii/0021999187900544`.

[290] S. Adamson, D. Kharlampidi, and A. Dementiev. “Stabilization of resonance states by an asymptotic Coulomb potential”. In: *The Journal of Chemical Physics* 128.2 (2008), p. 024101. DOI: `10.1063/1.2821102`. URL: `https://doi.org/10.1063/1.2821102`.

[291] Sergei Izrailev et al. "Steered Molecular Dynamics". In: *Proceedings of the 2nd International Symposium on Algorithms for Macromolecular Modelling*. Ed. by Peter Deuflhard et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 39–65. ISBN: 978-3-642-58360-5. DOI: `https://doi.org/10.1007/978-3-642-58360-5`.

[292] Johannes Kästner and Walter Thiel. "Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration"". In: *The Journal of Chemical Physics* 123.14 (2005), p. 144104. DOI: `10.1063/1.2052648`. URL: `https://doi.org/10.1063/1.2052648`.

[293] Shankar Kumar et al. "THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method". In: *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021. DOI: `https://doi.org/10.1002/jcc.540130812`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540130812`.

[294] James C. Gumbart, Benoît Roux, and Christophe Chipot. "Efficient Determination of Protein–Protein Standard Binding Free Energies from First Principles". In: *Journal of Chemical Theory and Computation* 9.8 (2013), pp. 3789–3798. DOI: `10.1021/ct400273t`. URL: `https://doi.org/10.1021/ct400273t`.

[295] Ozlem Ulucan and Volkhard Helms. "How Hydrophilic Proteins Form Non-specific Complexes". In: *The Journal of Physical Chemistry B* 119.33 (2015). PMID: 26218591, pp. 10524–10530. DOI: `10.1021/acs.jpcb.5b05831`. URL: `https://doi.org/10.1021/acs.jpcb.5b05831`.

[296] Donghyuk Suh et al. "String Method for Protein–Protein Binding Free-Energy Calculations". In: *Journal of Chemical Theory and Computation* 15.11 (2019). PMID: 31593627, pp. 5829–5844. DOI: `10.1021/acs.jctc.9b00499`. URL: `https://doi.org/10.1021/acs.jctc.9b00499`.

[297] Till Siebenmorgen and Martin Zacharias. "Evaluation of Predicted Protein–Protein Complexes by Binding Free Energy Simulations". In: *Journal of Chemical Theory and Computation* 15.3 (2019). PMID: 30698954, pp. 2071–2086. DOI: `10.1021/acs.jctc.8b01022`. URL: `https://doi.org/10.1021/acs.jctc.8b01022`.

[298] Charles H Bennett. "Efficient estimation of free energy differences from Monte Carlo data". In: *Journal of Computational Physics* 22.2 (1976), pp. 245 –268. ISSN: 0021-9991. DOI: `https://doi.org/10.1016/0021-9991(76)90078-4`. URL: `http://www.sciencedirect.com/science/article/pii/0021999176900784`.

[299] Michel A. Cuendet and Olivier Michielin. "Protein-Protein Interaction Investigated by Steered Molecular Dynamics: The TCR-pMHC Complex". In: *Biophysical Journal* 95.8 (2008), pp. 3575 –3590. ISSN: 0006-3495. DOI: `https://doi.org/10.1529/biophysj.108.131383`. URL: `http://www.sciencedirect.com/science/article/pii/S0006349508785023`.

[300] Roberto A. Rodriguez, Lili Yu, and Liao Y. Chen. "Computing Protein–Protein Association Affinity with Hybrid Steered Molecular Dynamics". In: *Journal of Chemical Theory and Computation* 11.9 (2015). PMID: 26366131, pp. 4427–4438. DOI: `10.1021/acs.jctc.5b00340`. URL: `https://doi.org/10.1021/acs.jctc.5b00340`.

[301] Alessandro Laio and Michele Parrinello. "Escaping free-energy minima". In: *Proceedings of the National Academy of Sciences* 99.20 (2002), pp. 12562–12566. ISSN: 0027-8424. DOI: `10.1073/pnas.202427399`. URL: `https://www.pnas.org/content/99/20/12562`.

[302] Alessandro Laio et al. "Assessing the Accuracy of Metadynamics". In: *The Journal of Physical Chemistry B* 109.14 (2005). PMID: 16851755, pp. 6714–6721. DOI: `10.1021/jp045424k`. URL: `https://doi.org/10.1021/jp045424k`.

[303] Vojtech Spiwok, Zoran Sucur, and Petr Hosek. "Enhanced sampling techniques in biomolecular simulations". In: *Biotechnology Advances* 33.6, Part 2 (2015). BioTech 2014 and 6th Czech-Swiss Biotechnology Symposium, pp. 1130 –1140. ISSN: 0734-9750. DOI: `https://doi.org/10.1016/j.biotechadv.2014.11.011`. URL: `http://www.sciencedirect.com/science/article/pii/S0734975014001864`.

[304] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. "Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method". In: *Phys. Rev. Lett.* 100 (2 Jan. 2008), p. 020603. DOI: `10.1103/PhysRevLett.100.020603`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.100.020603`.

[305] James F. Dama, Michele Parrinello, and Gregory A. Voth. "Well-Tempered Metadynamics Converges Asymptotically". In: *Phys. Rev. Lett.* 112 (24 June 2014), p. 240602. DOI: `10.1103/PhysRevLett.112.240602`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.112.240602`.

[306] Vittorio Limongelli, Massimiliano Bonomi, and Michele Parrinello. "Funnel metadynamics as accurate binding free-energy method". In: *Proceedings of the National Academy of Sciences* 110.16 (2013), pp. 6358–6363. ISSN: 0027-8424. DOI: `10.1073/pnas.1303186110`. URL: `https://www.pnas.org/content/110/16/6358`.

[307] Riccardo Capelli, Paolo Carloni, and Michele Parrinello. "Exhaustive Search of Ligand Binding Pathways via Volume-Based Metadynamics". In: *The Journal of Physical Chemistry Letters* 10.12 (2019), pp. 3495–3499. DOI: `10.1021/acs.jpclett.9b01183`. URL: `https://doi.org/10.1021/acs.jpclett.9b01183`.

[308] Paolo Raiteri et al. "Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics". In: *The Journal of Physical Chemistry B* 110.8 (2006). PMID: 16494409, pp. 3533–3539. DOI: `10.1021/jp054359r`. URL: `https://doi.org/10.1021/jp054359r`.

[309] Sadanand Singh, Chi-cheng Chiu, and Juan J. de Pablo. "Flux Tempered Metadynamics". In: *Journal of Statistical Physics* 145.4 (Nov. 2011), pp. 932–945. ISSN: 1572-9613. DOI: `10.1007/s10955-011-0301-0`. URL: `https://doi.org/10.1007/s10955-011-0301-0`.

[310] Omar Valsson, Pratyush Tiwary, and Michele Parrinello. "Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint". In: *Annual Review of Physical Chemistry* 67.1 (2016). PMID: 26980304, pp. 159–184. DOI: `10.1146/annurev-physchem-040215-112229`. URL: `https://doi.org/10.1146/annurev-physchem-040215-112229`.

[311] Pratyush Tiwary and Michele Parrinello. "From Metadynamics to Dynamics". In: *Phys. Rev. Lett.* 111 (23 Dec. 2013), p. 230602. DOI: `10.1103/PhysRevLett.111.230602`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.111.230602`.

[312] Matteo Salvalaglio, Pratyush Tiwary, and Michele Parrinello. "Assessing the Reliability of the Dynamics Reconstructed from Metadynamics". In: *Journal of Chemical Theory and Computation* 10.4 (2014). PMID: 26580360, pp. 1420–1425. DOI: `10.1021/ct500040r`. URL: `https://doi.org/10.1021/ct500040r`.

[313] Giovanni Bussi et al. "Free-Energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics". In: *Journal of the American Chemical Society* 128.41 (2006). PMID: 17031956, pp. 13435–13441. DOI: `10.1021/ja062463w`. URL: `https://doi.org/10.1021/ja062463w`.

[314] Alessandro Barducci et al. "Free-energy landscape of protein oligomerization from atomistic simulations". In: *Proceedings of the National Academy of Sciences* 110.49 (2013), E4708–E4713. ISSN: 0027-8424. DOI: `10.1073/pnas.1320077110`. URL: `https://www.pnas.org/content/110/49/E4708`.

[315] Puja Banerjee, Sayantan Mondal, and Biman Bagchi. "Insulin dimer dissociation in aqueous solution: A computational study of free energy landscape and evolving microscopic structure along the reaction pathway". In: *The Journal of Chemical Physics* 149.11 (2018), p. 114902. DOI: `10.1063/1.5042290`. URL: `https://doi.org/10.1063/1.5042290`.

[316] V. Lindahl, J. Lidmar, and B. Hess. "Accelerated weight histogram method for exploring free energy landscapes". In: *The Journal of Chemical Physics* 141.4 (2014), p. 044110. DOI: `10.1063/1.4890371`. URL: `https://doi.org/10.1063/1.4890371`.

[317] Massimiliano Bonomi et al. "PLUMED: A portable plugin for free-energy calculations with molecular dynamics". In: *Computer Physics Communications* 180.10 (2009), pp. 1961 –1972. ISSN: 0010-4655. DOI: `https://doi.org/10.1016/j.cpc.2009.05.011`. URL: `http://www.sciencedirect.com/science/article/pii/S001046550900157X`.

[318] Gareth A. Tribello et al. "PLUMED 2: New feathers for an old bird". In: *Computer Physics Communications* 185.2 (2014), pp. 604 –613. ISSN: 0010-4655. DOI: `https://doi.org/10.1016/j.cpc.2013.09.018`. URL: `http://www.sciencedirect.com/science/article/pii/S0010465513003196`.

[319] Melina Mottin, Paulo C. T. Souza, and Munir S. Skaf. "Molecular Recognition of PPARγ by Kinase Cdk5/p25: Insights from a Combination of Protein–Protein Docking and Adaptive Biasing Force Simulations". In: *The Journal of Physical Chemistry B* 119.26 (2015). PMID: 26047365, pp. 8330–8339. DOI: `10.1021/acs.jpcb.5b04269`. URL: `https://doi.org/10.1021/acs.jpcb.5b04269`.

[320] Ryuhei Harada and Akio Kitao. "Parallel cascade selection molecular dynamics (PaCS-MD) to generate conformational transition pathway". In: *The Journal of Chemical Physics* 139.3 (2013), p. 035103. DOI: `10.1063/1.4813023`. URL: `https://doi.org/10.1063/1.4813023`.

[321] Duy Phuoc Tran et al. "Protein-Ligand Dissociation Simulated by Parallel Cascade Selection Molecular Dynamics". In: *Journal of Chemical Theory and Computation* 14.1 (2018). PMID: 29182324, pp. 404–417. DOI: `10.1021/acs.jctc.7b00504`. URL: `https://doi.org/10.1021/acs.jctc.7b00504`.

[322] Duy Phuoc Tran and Akio Kitao. "Kinetic Selection and Relaxation of the Intrinsically Disordered Region of a Protein upon Binding". In: *Journal of Chemical Theory and Computation* 16.4 (2020). PMID: 32192337, pp. 2835–2845. DOI: `10.1021/acs.jctc.9b01203`. URL: `https://doi.org/10.1021/acs.jctc.9b01203`.

[323] Yuji Sugita and Yuko Okamoto. "Replica-exchange molecular dynamics method for protein folding". In: *Chemical Physics Letters* 314.1 (1999), pp. 141 –151. ISSN: 0009-2614. DOI: `https://doi.org/10.1016/S0009-2614(99)01123-9`. URL: `http://www.sciencedirect.com/science/article/pii/S0009261499011239`.

[324] E Marinari and G Parisi. "Simulated Tempering: A New Monte Carlo Scheme". In: *Europhysics Letters (EPL)* 19.6 (July 1992), pp. 451–458. DOI: `10.1209/0295-5075/19/6/002`. URL: `https://doi.org/10.1209/0295-5075/19/6/002`.

[325] David J. Earl and Michael W. Deem. "Parallel tempering: Theory, applications, and new perspectives". In: *Phys. Chem. Chem. Phys.* 7 (23 2005), pp. 3910–3916. DOI: `10.1039/B509983H`. URL: `http://dx.doi.org/10.1039/B509983H`.

[326] Yuji Sugita and Yuko Okamoto. "Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape". In: *Chemical Physics Letters* 329.3 (2000), pp. 261 –270. ISSN: 0009-2614. DOI: `https://doi.org/10.1016/S0009-2614(00)00999-4`. URL: `http://www.sciencedirect.com/science/article/pii/S0009261400009994`.

[327] Dari Kimanius et al. "SAXS-Guided Metadynamics". In: *Journal of Chemical Theory and Computation* 11.7 (2015). PMID: 26575782, pp. 3491–3498. DOI: `10.1021/acs.jctc.5b00299`. URL: `https://doi.org/10.1021/acs.jctc.5b00299`.

[328] Gregory R. Bowman, Vijay S. Pande, and Frank Noé. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*. Vol. 797. Advances in Experimental Medicine and Biology. Springer Netherlands, 2014. ISBN: 978-94-007-7606-7. DOI: `10.1007/978-94-007-7606-7`.

[329] Robert T. McGibbon and Vijay S. Pande. "Variational cross-validation of slow dynamical modes in molecular kinetics". In: *The Journal of Chemical Physics* 142.12 (2015), p. 124105. DOI: `10.1063/1.4916292`. URL: `https://doi.org/10.1063/1.4916292`.

[330] Martin K. Scherer et al. "Variational selection of features for molecular kinetics". In: *The Journal of Chemical Physics* 150.19 (2019), p. 194108. DOI: `10.1063/1.5083040`. URL: `https://doi.org/10.1063/1.5083040`.

[331] Hao Wu and Frank Noé. "Variational Approach for Learning Markov Processes from Time Series Data". In: *Journal of Nonlinear Science* 30.1 (Feb. 2020), pp. 23–66. ISSN: 1432-1467. DOI: `10.1007/s00332-019-09567-y`. URL: `https://doi.org/10.1007/s00332-019-09567-y`.

[332] Yunhui Ge and Vincent A. Voelz. "Markov State Models to Elucidate Ligand Binding Mechanism". In: *Protein-Ligand Interactions and Drug Design*. Ed. by Flavio Ballante. New York, NY: Springer US, 2021, pp. 239–259. ISBN: 978-1-0716-1209-5. DOI: `10.1007/978-1-0716-1209-5_14`. URL: `https://doi.org/10.1007/978-1-0716-1209-5_14`.

[333] J. B. MacQueen. "Some Methods for Classification and Analysis of Multi-Variate Observations". In: *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*. Ed. by L. M. Le Cam and J. Neyman. Vol. 1. University of California Press, 1967, pp. 281–297.

[334] Gregory R. Bowman et al. "Progress and challenges in the automated construction of Markov state models for full protein systems". In: *The Journal of Chemical Physics* 131.12 (2009), p. 124101. DOI: `10.1063/1.3216567`. URL: `https://doi.org/10.1063/1.3216567`.

[335] Marcus Weber and Tobias Galliat. *Characterization of Transition States in Conformational Dynamics using Fuzzy Sets*. eng. Tech. rep. 02-12. Takustr. 7, 14195 Berlin: ZIB, 2002.

[336] Peter Deuflhard and Marcus Weber. "Robust Perron cluster analysis in conformation dynamics". In: *Linear Algebra and its Applications* 398 (2005). Special Issue on Matrices and Mathematical Biology, pp. 161–184. ISSN: 0024-3795. DOI: `https://doi.org/10.1016/j.laa.2004.10.026`. URL: `https://www.sciencedirect.com/science/article/pii/S0024379504004689`.

[337] Susanna Röblitz and Marcus Weber. "Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification". In: *Advances in Data Analysis and Classification* 7.2 (June 2013), pp. 147–179. DOI: `10.1007/s11634-013-0134-6`. URL: `https://doi.org/10.1007/s11634-013-0134-6`.

[338] A. J. Muslin et al. "Interaction of 14-3-3 with signaling proteins is mediated by the recognition of phosphoserine". In: *Cell* 84.6 (1996), pp. 889–897. ISSN: 00928674. DOI: `10.1016/S0092-8674(00)81067-3`.

[339] Brenda Kostelecky et al. "Recognition of an intra-chain tandem 14-3-3 binding site within PKCϵ". In: *EMBO Reports* 10.9 (Sept. 2009), pp. 983–989. ISSN: 1469221X. DOI: `10.1038/embor.2009.150`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1038/embor.2009.150`.

[340] Michael B. Yaffe. "How do 14-3-3 proteins work? - Gatekeeper phosphorylation and the molecular anvil hypothesis". In: *FEBS Letters* 513.1 (2002), pp. 53–57. ISSN: 00145793. DOI: `10.1016/S0014-5793(01)03288-4`.

[341] Manuela Molzan and Christian Ottmann. "Synergistic Binding of the Phosphorylated S233- and S259-Binding Sites of C-RAF to One 14-3-3ζ Dimer". In: *Journal of Molecular Biology* 423.4 (2012), pp. 486 –495. ISSN: 0022-2836. DOI: `https://doi.org/10.1016/j.jmb.2012.08.009`. URL: `http://www.sciencedirect.com/science/article/pii/S0022283612006717`.

[342] Xiaowen Yang et al. "Structural basis for protein–protein interactions in the 14-3-3 protein family". In: *Proceedings of the National Academy of Sciences* 103.46 (2006), pp. 17237–17242. ISSN: 0027-8424. DOI: `10.1073/pnas.0605779103`. URL: `https://www.pnas.org/content/103/46/17237`.

[343] Eric F. Pettersen et al. "UCSF Chimera - A visualization system for exploratory research and analysis". In: *Journal of Computational Chemistry* 25.13 (2004), pp. 1605–1612. DOI: `10.1002/jcc.20084`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20084`.

[344] Katrin Rittinger et al. "Structural analysis of 14-3-3 phosphopeptide complexes identifies a dual role for the nuclear export signal of 14-3-3 in ligand binding". In: *Molecular Cell* 4.2 (1999), pp. 153–166. ISSN: 10972765. DOI: `10.1016/S1097-2765(00)80363-9`.

[345] Roland L Dunbrack. "Rotamer Libraries in the 21st Century". In: *Current Opinion in Structural Biology* 12.4 (2002), pp. 431 –440. ISSN: 0959-440X. DOI: `https://doi.org/10.1016/S0959-440X(02)00344-5`. URL: `http://www.sciencedirect.com/science/article/pii/S0959440X02003445`.

[346] Vytautas Gapsys and Bert L. De Groot. "pmx Webserver: A User Friendly Interface for Alchemistry". In: *Journal of Chemical Information and Modeling* 57.2 (2017), pp. 109–114. ISSN: 15205142. DOI: `10.1021/acs.jcim.6b00498`. URL: `http://pmx.mpibpc.mpg.de`.

[347] Berk Hess et al. "LINCS: A linear constraint solver for molecular simulations". In: *Journal of Computational Chemistry* 18.12 (1997), pp. 1463–1472. DOI: `10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/\%28SICI\%291096-987X\%28199709\%2918\%3A12\%3C1463\%3A\%3AAID-JCC4\%3E3.0.CO\%3B2-H`.

[348] Shuichi Nosé and M.L. Klein. "Constant pressure molecular dynamics for molecular systems". In: *Molecular Physics* 50.5 (1983), pp. 1055–1076. DOI: `10.1080/00268978300102851`. URL: `https://doi.org/10.1080/00268978300102851`.

[349] Robert T. McGibbon et al. "MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories". In: *Biophysical Journal* 109.8 (2015), pp. 1528 –1532. DOI: `10.1016/j.bpj.2015.08.015`.

[350] E.N. Baker and R.E. Hubbard. "Hydrogen bonding in globular proteins". In: *Progress in Biophysics and Molecular Biology* 44.2 (1984), pp. 97 –179. ISSN: 0079-6107. DOI: `https://doi.org/10.1016/0079-6107(84)90007-5`. URL: `http://www.sciencedirect.com/science/article/pii/0079610784900075`.

[351] Maik Goette and Helmut Grubmüller. "Accuracy and convergence of free energy differences calculated from nonequilibrium switching processes". In: *Journal of Computational Chemistry* 30.3 (2009), pp. 447–456. DOI: `10.1002/jcc.21073`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.21073`.

[352] Mika A. Kastenholz and Philippe H. Hünenberger. "Computation of methodology-independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation". In: *The Journal of Chemical Physics* 124.22 (2006), p. 224501. DOI: `10.1063/1.2201698`. URL: `https://doi.org/10.1063/1.2201698`.

[353] Maria M. Reif and Philippe H. Hünenberger. "Computation of methodology-independent single-ion solvation properties from molecular simulations. III. Correction terms for the solvation free energies, enthalpies, entropies, heat capacities, volumes, compressibilities, and expansivities of solvated ions". In: *The Journal of Chemical Physics* 134.14 (2011), p. 144103. DOI: `10.1063/1.3567020`. URL: `https://doi.org/10.1063/1.3567020`.

[354] Wei Chen et al. "Accurate Calculation of Relative Binding Free Energies between Ligands with Different Net Charges". In: *Journal of Chemical Theory and Computation* 14.12 (2018), pp. 6346–6358. ISSN: 15499626. DOI: `10.1021/acs.jctc.8b00825`. URL: `https://pubs.acs.org/sharingguidelines`.

[355] Hui Xiong et al. "Free energy calculations with non-equilibrium methods: Applications of the Jarzynski relationship". In: *Theoretical Chemistry Accounts* 116.1-3 (2006), pp. 338–346. ISSN: 1432881X. DOI: `10.1007/s00214-005-0072-2`.

[356] Yong Xie, Yanan Jiang, and Dor Ben-Amotz. "Detection of amino acid and peptide phosphate protonation using Raman spectroscopy". In: *ANALYTICAL BIOCHEMISTRY* 343 (2005), pp. 223–230. DOI: `10.1016/j.ab.2005.05.038`. URL: `www.elsevier.com/locate/yabio`.

[357] Onur Serçinoğlu and Pemra Ozbek. "gRINN: a tool for calculation of residue interaction energies and protein energy network analysis of molecular dynamics simulations". In: *Nucleic Acids Research* 46.W1 (May 2018), W554–W562. ISSN: 0305-1048. DOI: `10.1093/nar/gky381`. URL: `https://doi.org/10.1093/nar/gky381`.

[358] Yifei Kong and Martin Karplus. "Signaling pathways of PDZ2 domain: A molecular dynamics interaction correlation analysis". In: *Proteins: Structure, Function and Bioinformatics* 74.1 (2009), pp. 145–154. ISSN: 08873585. DOI: `10.1002/prot.22139`.

[359] Giulia Morra, Alessandro Genoni, and Giorgio Colombo. "Mechanisms of Differential Allosteric Modulation in Homologous Proteins: Insights from the Analysis of Internal Dynamics and Energetics of PDZ Domains". In: *Journal of Chemical Theory and Computation* 10.12 (2014). PMID: 26583250, pp. 5677–5689. DOI: `10.1021/ct500326g`. URL: `https://doi.org/10.1021/ct500326g`.

[360] Nicolas Panel et al. "Accurate PDZ/Peptide Binding Specificity with Additive and Polarizable Free Energy Simulations". In: *Biophysical Journal* 114.5 (2018), pp. 1091–1102. ISSN: 0006-3495. DOI: `doi.org/10.1016/j.bpj.2018.01.008`. URL: `https://www.sciencedirect.com/science/article/pii/S000634951830105X`.

[361] Mahalakshmi Harish et al. "A Novel Determinant of PSMD9 PDZ Binding Guides the Evolution of the First Generation of Super Binding Peptides". In: *Biochemistry* 58.32 (2019). PMID: 31287951, pp. 3422–3433. DOI: `10.1021/acs.biochem.9b00308`. URL: `https://doi.org/10.1021/acs.biochem.9b00308`.

[362] Sebastian Charbonnier et al. "The Structural and Dynamic Response of MAGI-1 PDZ1 with Noncanonical Domain Boundaries to the Binding of Human Papillomavirus E6". In: *Journal of Molecular Biology* 406.5 (2011), pp. 745–763. ISSN: 0022-2836. DOI: `https://doi.org/10.1016/j.jmb.2011.01.015`. URL: `https://www.sciencedirect.com/science/article/pii/S0022283611000325`.

[363] Jun Zhang et al. "Crystallographic and Nuclear Magnetic Resonance Evaluation of the Impact of Peptide Binding to the Second PDZ Domain of Protein Tyrosine Phosphatase 1E". In: *Biochemistry* 49.43 (2010), pp. 9280–9291. DOI: `10.1021/bi101131f`. URL: `https://doi.org/10.1021/bi101131f`.

[364] Chresten R. Søndergaard et al. "Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values". In: *Journal of Chemical Theory and Computation* 7.7 (2011), pp. 2284–2295. DOI: `10.1021/ct200133y`. URL: `https://doi.org/10.1021/ct200133y`.

[365] Mats H. M. Olsson et al. "PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions". In: *Journal of Chemical Theory and Computation* 7.2 (2011), pp. 525–537. DOI: `10.1021/ct100578z`. URL: `https://doi.org/10.1021/ct100578z`.

[366] J. Andrew McCammon et al. "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations". In: *Nucleic Acids Research* 32 (July 2004), W665–W667. ISSN: 0305-1048. DOI: `10.1093/nar/gkh381`. URL: `https://dx.doi.org/10.1093/nar/gkh381`.

[367] Martin K. Scherer et al. "PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models". In: *Journal of Chemical Theory and Computation* 11 (Oct. 2015), pp. 5525–5542. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.5b00743`. URL: `http://dx.doi.org/10.1021/acs.jctc.5b00743` (visited on 10/19/2015).

[368] J. MacQueen. "Some Methods for Classification and Analysis of Multivariate Observations". In: *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability - Vol. 1*. Ed. by L. M. Le Cam and J. Neyman. University of California Press, Berkeley, USA, 1967, pp. 281–297.

[369] Wolfgang Kabsch and Christian Sander. "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features". In: *Biopolymers* 22.12 (1983), pp. 2577–2637. DOI: `10.1002/bip.360221211`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.360221211`.

[370] Jiří Vymětal, Veronika Jurásková, and Jiří Vondrášek. "AMBER and CHARMM Force Fields Inconsistently Portray the Microscopic Details of Phosphorylation". In: *Journal of Chemical Theory and Computation* 15.1 (2019), pp. 665–679. DOI: `10.1021/acs.jctc.8b00715`. URL: `https://doi.org/10.1021/acs.jctc.8b00715`.

[371] Dolors Grillo-Bosch, Daniel Choquet, and Matthieu Sainlos. "Inhibition of PDZ domain-mediated interactions". In: *Drug Discovery Today: Technologies* 10.4 (2013), e531–e540. ISSN: 1740-6749. DOI: `10.1016/j.ddtec.2012.10.003`. URL: `https://www.sciencedirect.com/science/article/pii/S1740674912000790`.

[372] Kumlesh K. Dev. "Making protein interactions druggable: targeting PDZ domains". In: *Nature Reviews Drug Discovery* 3.12 (Dec. 2004), pp. 1047–1056. ISSN: 1474-1784. DOI: `10.1038/nrd1578`. URL: `https://doi.org/10.1038/nrd1578`.

[373] Naoaki Fujii et al. "Rational design of a nonpeptide general chemical scaffold for reversible inhibition of PDZ domain interactions". In: *Bioorganic & Medicinal Chemistry Letters* 17.2 (2007), pp. 549–552. ISSN: 0960-894X. DOI: `doi.org/10.1016/j.bmcl.2006.10.006`. URL: `https://www.sciencedirect.com/science/article/pii/S0960894X06011747`.

[374] David Grandy et al. "Discovery and Characterization of a Small Molecule Inhibitor of the PDZ Domain of Dishevelled". In: *Journal of Biological Chemistry* 284.24 (June 2009), pp. 16256–16263. ISSN: 0021-9258. DOI: `10.1074/jbc.M109.009647`. URL: `https://doi.org/10.1074/jbc.M109.009647`.

[375] Jufang Shan et al. "Synthesis of Potent Dishevelled PDZ Domain Inhibitors Guided by Virtual Screening and NMR Studies". In: *Chemical Biology & Drug Design* 79.4 (2012), pp. 376–383. DOI: `doi.org/10.1111/j.1747-0285.2011.01295.x`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1747-0285.2011.01295.x`.

[376] Shuangzhe Zhang et al. "Imaging and Inhibiting: A Dual Function Molecular Flare for Cancer Cells". In: *Analytical Chemistry* 91.21 (2019). PMID: 31571476, pp. 13501–13507. DOI: `10.1021/acs.analchem.9b02558`. URL: `https://doi.org/10.1021/acs.analchem.9b02558`.

[377] Nikolaj R Christensen et al. "A high-affinity, bivalent PDZ domain inhibitor complexes PICK1 to alleviate neuropathic pain". In: *EMBO Molecular Medicine* 12.6 (2020), e11248. DOI: `doi.org/10.15252/emmm.201911248`. URL: `https://www.embopress.org/doi/abs/10.15252/emmm.201911248`.

[378] Michael J. Thun, Mohan M. Namboodiri, and Clark W. Heath. "Aspirin Use and Reduced Risk of Fatal Colon Cancer". In: *New England Journal of Medicine* 325.23 (1991). PMID: 1669840, pp. 1593–1596. DOI: `10.1056/NEJM199112053252301`. URL: `https://doi.org/10.1056/NEJM199112053252301`.

[379] Lynn Rosenberg et al. "A Hypothesis: Nonsteroidal Anti-Inflammatory Drugs Reduce the Incidence of Large-Bowel Cancer". In: *JNCI: Journal of the National Cancer Institute* 83.5 (Mar. 1991), pp. 355–358. ISSN: 0027-8874. DOI: `10.1093/jnci/83.5.355`. URL: `https://doi.org/10.1093/jnci/83.5.355`.

[380] Okhee Suh, Curtis Mettlin, and Nicholas J. Petrelli. "Aspirin use, cancer, and polyps of the large bowel". In: *Cancer* 72.4 (1993), pp. 1171–1177. DOI: `https://doi.org/10.1002/1097-0142(19930815)72:4<1171::AID-CNCR2820720407>3.0.CO;2-D`. URL: `https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.1002/1097-0142%2819930815%2972%3A4%3C1171%3A%3AAID-CNCR2820720407%3E3.0.CO%3B2-D`.

[381] Edward Giovannucci et al. "Aspirin Use and the Risk for Colorectal Cancer and Adenoma in Male Health Professionals". In: *Annals of Internal Medicine* 121.4 (1994). PMID: 8037405, pp. 241–246. DOI: `10.7326/0003-4819-121-4-199408150-00001`. URL: `https://www.acpjournals.org/doi/abs/10.7326/0003-4819-121-4-199408150-00001`.

[382] Edward Giovannucci et al. "Aspirin and the Risk of Colorectal Cancer in Women". In: *New England Journal of Medicine* 333.10 (1995). PMID: 7637720, pp. 609–614. DOI: `10.1056/NEJM199509073331001`. URL: `https://doi.org/10.1056/NEJM199509073331001`.

[383] Yumi Yamamoto et al. "Sulindac Inhibits Activation of the NF-κB Pathway *". In: *Journal of Biological Chemistry* 274.38 (Sept. 1999), pp. 27307–27314. ISSN: 0021-9258. DOI: `10.1074/jbc.274.38.27307`. URL: `https://doi.org/10.1074/jbc.274.38.27307`.

[384] Nick Barker and Hans Clevers. "Mining the Wnt pathway for cancer therapeutics". In: *Nature Reviews Drug Discovery* 5.12 (Dec. 2006), pp. 997–1014. ISSN: 1474-1784. DOI: `10.1038/nrd2154`. URL: `https://doi.org/10.1038/nrd2154`.

[385] Benjamin N.R. Cheyette et al. "Dapper, a Dishevelled-Associated Antagonist of β-Catenin and JNK Signaling, Is Required for Notochord Formation". In: *Developmental Cell* 2.4 (2002), pp. 449–461. ISSN: 1534-5807. DOI: `https://doi.org/10.1016/S1534-5807(02)00140-5`. URL: `https://www.sciencedirect.com/science/article/pii/S1534580702001405`.

[386] K. Vanommeslaeghe and A. D. MacKerell. "Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing". In: *Journal of Chemical Information and Modeling* 52.12 (2012). PMID: 23146088, pp. 3144–3154. DOI: `10.1021/ci300363c`. URL: `https://doi.org/10.1021/ci300363c`.

[387] K. Vanommeslaeghe, E. Prabhu Raman, and A. D. MacKerell. "Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges". In: *Journal of Chemical Information and Modeling* 52.12 (2012). PMID: 23145473, pp. 3155–3168. DOI: `10.1021/ci3003649`. URL: `https://doi.org/10.1021/ci3003649`.

[388] Anmol Kumar, Ozge Yoluk, and Alexander D. MacKerell Jr. “FFParam: Standalone package for CHARMM additive and Drude polarizable force field parametrization of small molecules”. In: *Journal of Computational Chemistry* 41.9 (2020), pp. 958–970. DOI: `https://doi.org/10.1002/jcc.26138`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.26138`.

[389] Peter Eastman et al. “OpenMM 7: Rapid development of high performance algorithms for molecular dynamics”. In: *PLOS Computational Biology* 13.7 (July 2017), pp. 1–17. DOI: `10.1371/journal.pcbi.1005659`. URL: `https://doi.org/10.1371/journal.pcbi.1005659`.

[390] Daniel G. A. Smith et al. “PSI4 1.4: Open-source software for high-throughput quantum chemistry”. In: *The Journal of Chemical Physics* 152.18 (2020), p. 184108. DOI: `10.1063/5.0006002`. URL: `https://doi.org/10.1063/5.0006002`.

[391] R. Ditchfield, W. J. Hehre, and J. A. Pople. “Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules”. In: *The Journal of Chemical Physics* 54.2 (1971), pp. 724–728. DOI: `10.1063/1.1674902`. URL: `https://doi.org/10.1063/1.1674902`.

[392] *CHARMM Force Field Files*. URL: `https://mackerell.umaryland.edu/charmm_ff.shtml`.

[393] James A. Maier et al. “ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB”. In: *Journal of Chemical Theory and Computation* 11.8 (2015). PMID: 26574453, pp. 3696–3713. DOI: `10.1021/acs.jctc.5b00255`. URL: `https://doi.org/10.1021/acs.jctc.5b00255`.

[394] Chuan Tian et al. “ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution”. In: *Journal of Chemical Theory and Computation* 16.1 (2020). PMID: 31714766, pp. 528–552. DOI: `10.1021/acs.jctc.9b00591`. URL: `https://doi.org/10.1021/acs.jctc.9b00591`.

[395] *Absolute Binding Free Energy - Gromacs 2016*. Sept. 2021. URL: `http://alchemistry.org/wiki/Absolute_Binding_Free_Energy_-_Gromacs_2016`.