Understanding disordered and membrane protein recognition by molecular dynamics
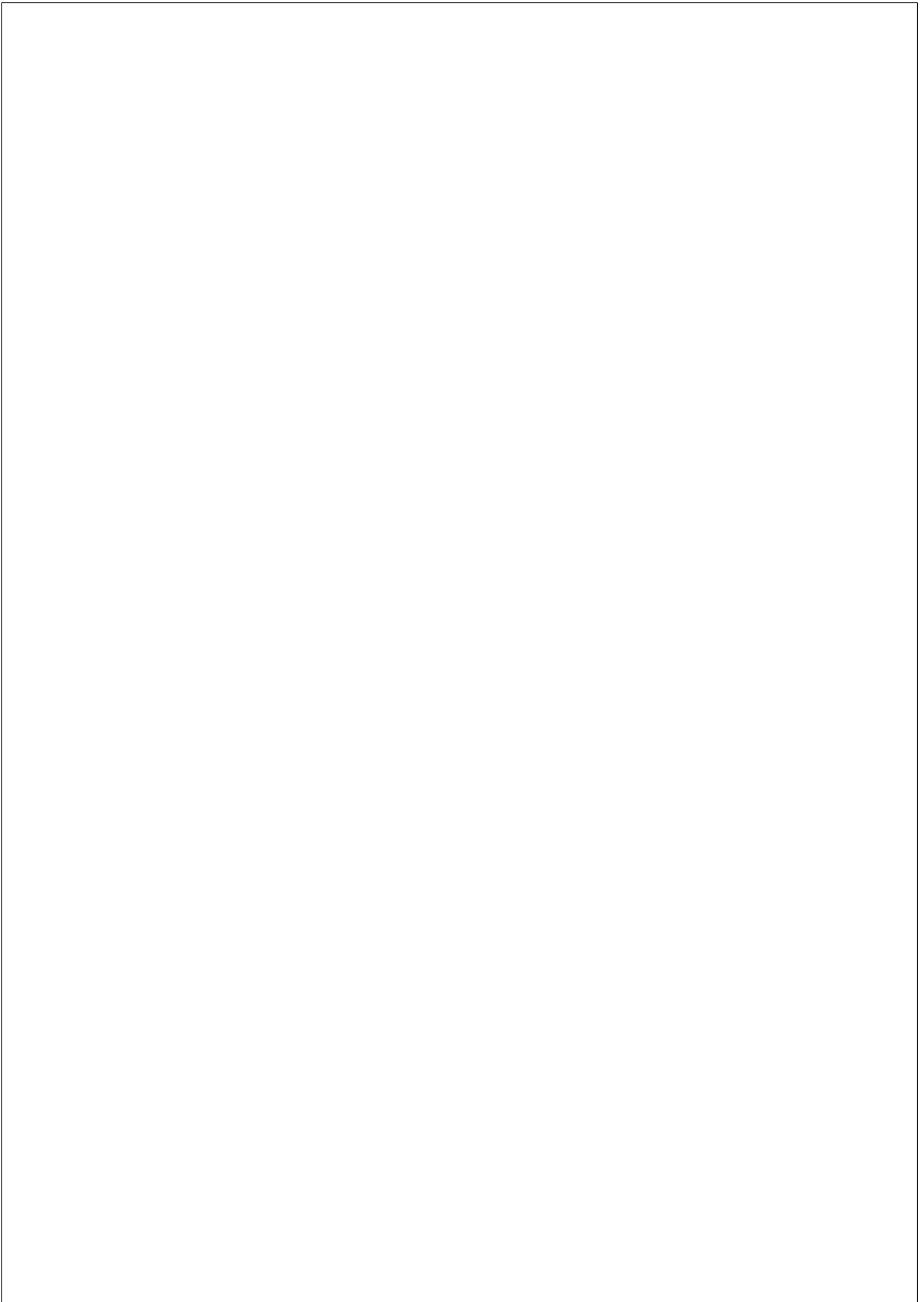
Nathaniel H. Stanley

TESI DOCTORAL UPF / ANY 2015

DIRECTOR DE LA TESI
Dr. Gianni De Fabritiis
Departament de Cièncias Experimentals i de la Salut

upf. Universitat Pompeu Fabra
Barcelona

Dedicated to the people I've watched die bad deaths. Hopefully, either with science or with ethics, we can avoid that someday.

Acknowledgements

I would like to thank the volunteers of GPUGRID.net and Donate@Home for donating their time, hardware, bitcoins, and other resources to support my research. Without them, my research would not have been possible. I would like to thank my labmates for both assisting me and putting up with me. I would like to thank my family, who for many years have given me their support and dealt with my absence.
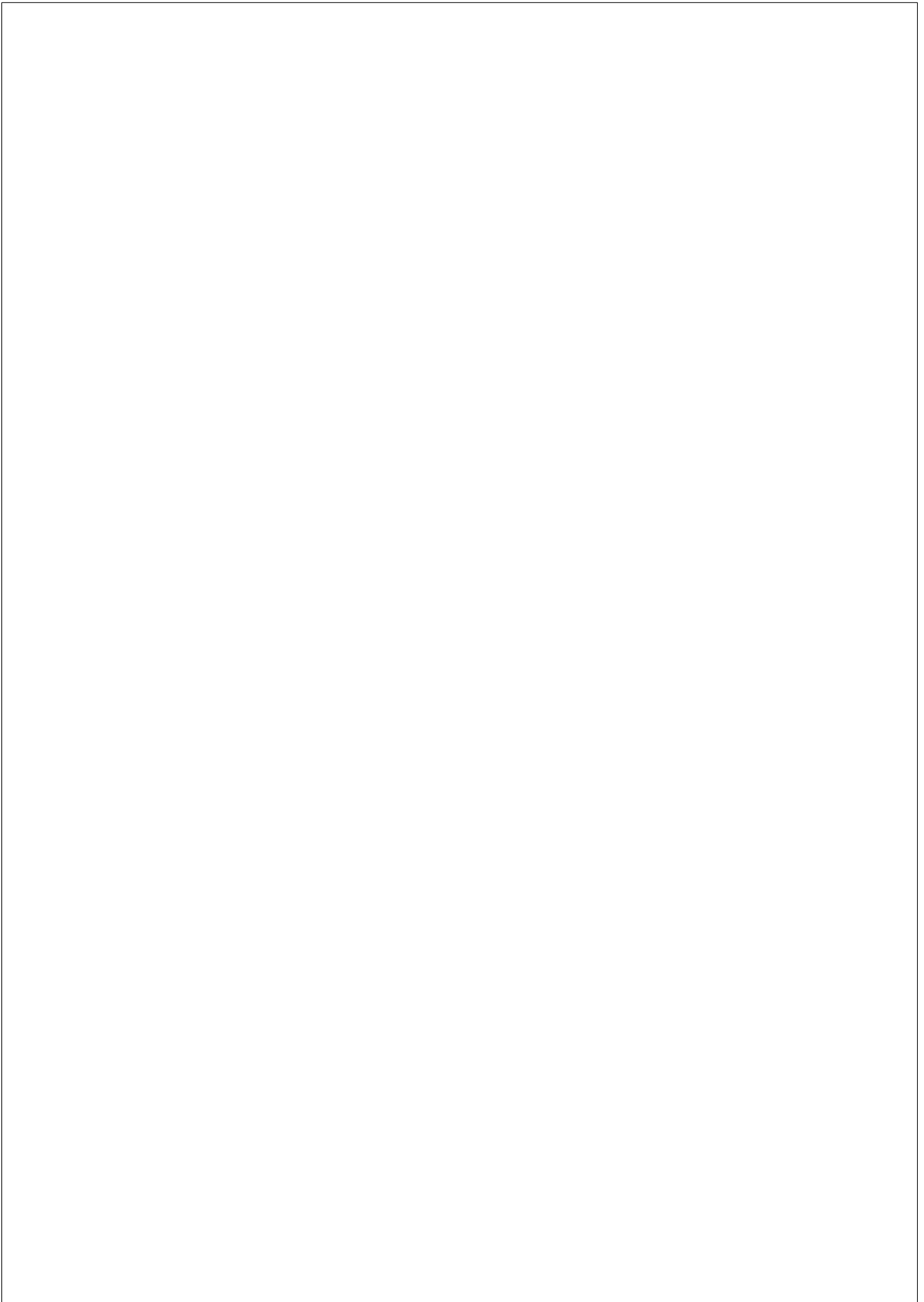
Abstract

All biological processes are governed by interactions of macromolecules that occur at atomic scale. However, our ability to directly observe such processes is often limited by experimental constraints due to the characteristic scales at which they occur. Such limitations mandate the use of modeling techniques such as molecular dynamics simulations to extend our understanding of these phenomena. The goal of this thesis has been to use molecular dynamics simulations, in conjunction with advanced analysis techniques, to elucidate biological processes at the atomistic scale. We have used the distributed computing project GPUGRID.net and Markov State Model analysis to study molecular processes in disordered proteins and membranes systems. In each case we have been able to give a full atomic picture of events only hinted at by other methods, and in some cases we observe things entirely hidden from other methods. These successes reinforce the importance of molecular simulations as an exploratory tool in the biological sciences.
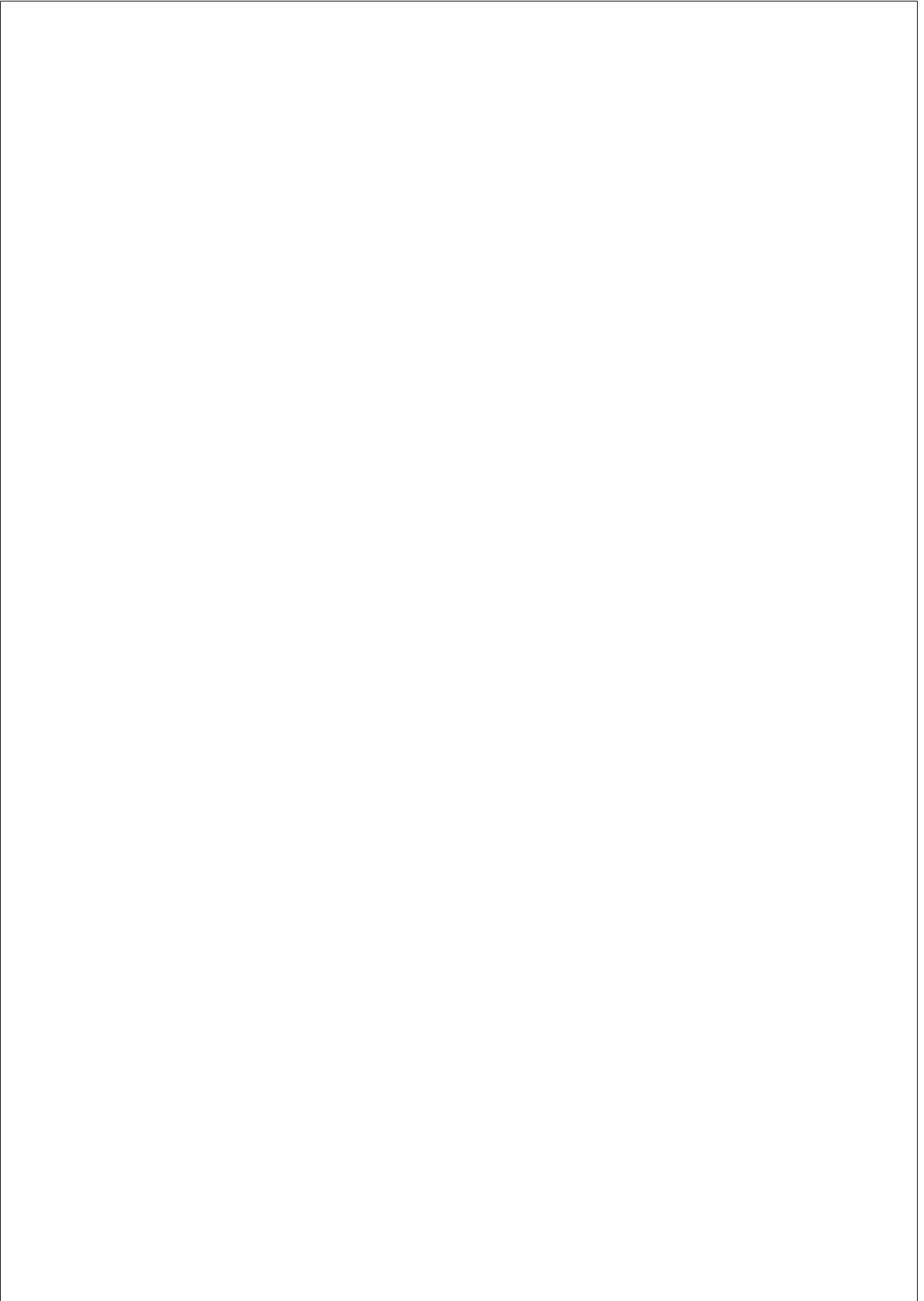
Resum

Tots els processos biològics estan governats per interaccions de macro-molècules que ocorren a escala atòmica. No obstant, la nostra capacitat d'observar directament aquests processos sovint està limitada per restriccions experimentals degut a les escales característiques en que ocurreixen. Tals limitacions requereixen l'ús de tècniques de modelatge com simulacions de dinàmica molecular que permeten extendre la nostra comprensió d'aquests fenòmens. L'objectiu d'aquesta tesi ha estat aplicar simulacions de dinàmica molecular, conjuntament amb tècniques d'anàlisi avançada, per dilucidar processos biològics a escala atòmica. Hem utilitzat el projecte de computaciò distribuïda GPUGRID.net i l'anàlisi mitjançant Models d'Estat de Markov per estudiar processos moleculars en sistemes de membranes i proteïnes desordenades. En alguns casos hem estat capaços de retratar fenòmens amb indicis proporcionats per altres mètodes i en d'altres n'hem pogut observar de totalment ocults a altres mètodes. Aquests èxits reforcen la importància de les simulacions moleculars com a eina exploratòria en les ciències biològiques.

Preface

One of the fundamental obstacles—and accordingly one of the main lessons—of the scientific process is that we are always working on a "black box" problem. In our pursuit of fundamental truths, be they in the macroscopic or microscopic world, we never have direct access to the thing we want to know. Even for something "simple" like the law of gravity, observations of physical objects had to be made before we could dispel wrong intuitions and write the equation down. And the equation itself is just an abstraction of something much more complex. Empirical, repeatable, falsifiable experiments are an inextricable part of the scientific process, because they are our only way to find out what's inside the black box.

The goal of a PhD program is to train ourselves to prod at that black box. We learn what kinds of questions we need to ask ourselves before we even begin (What do we really know? What do we need to know?). We learn what tools are best for answering different kinds of questions and how to use them. And we learn to be properly skeptical of what comes out of our experiments and other's experiments.

Like many theses before this one, this thesis is essentially the abridged version of that struggle for me. I have attempted to improve our understanding of protein motions and their interactions with other biomolecules, suffering a lot of failures and frustrations as well as a good amount of luck and success. The successes are stressed here, because that's how we handle these things. But in reality we learn most from our struggles and failures, and I hope that my struggles are somehow apparent between the lines.
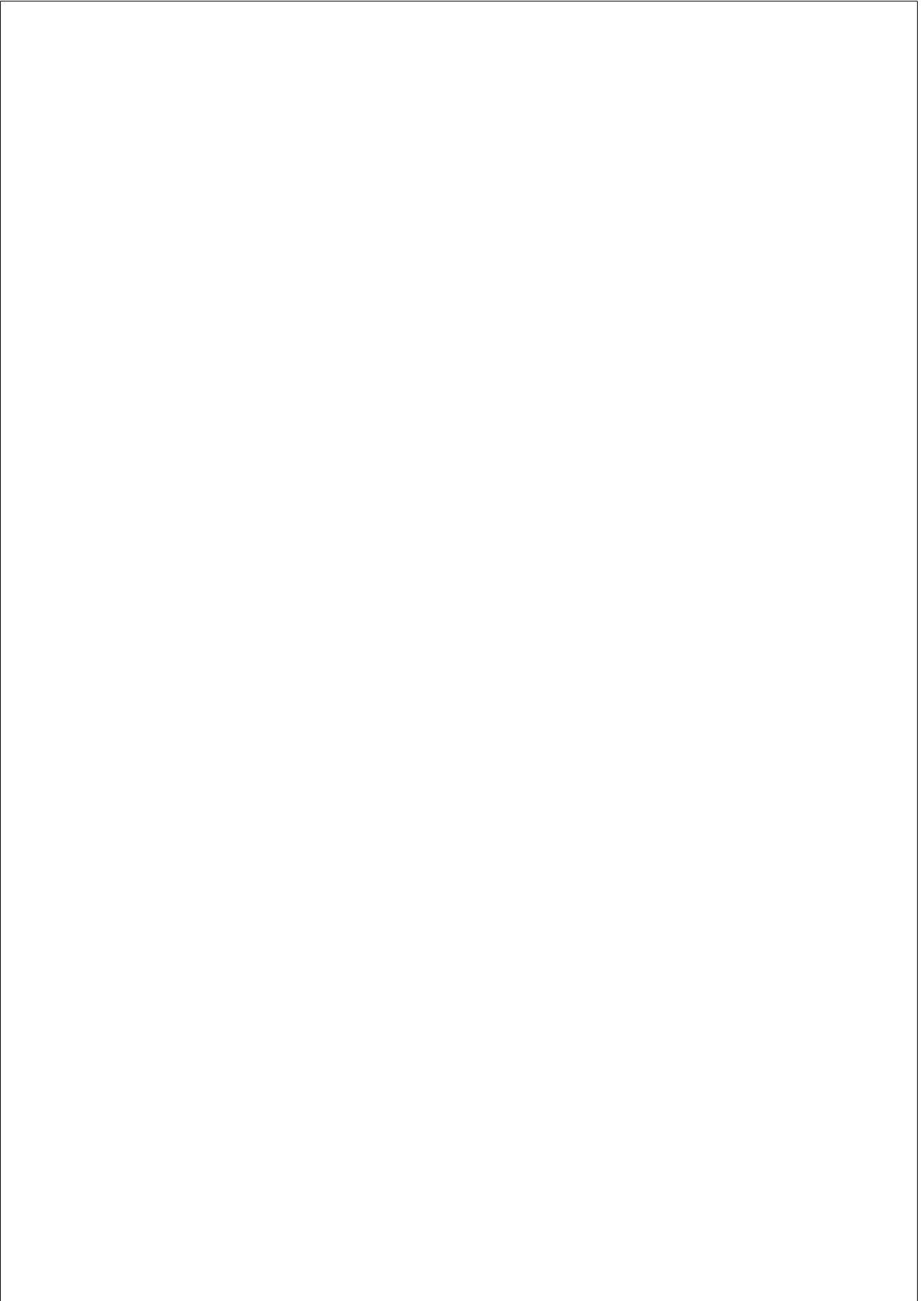
Contents

List of Figures

Chapter 1

INTRODUCTION

1.1 Peering at Atoms

All biology is dictated by dynamic, atomic interactions of a vast array of molecules. In order to understand biological systems and cure diseases, we need to be able to understand the behavior and interactions of these molecules. In a convenient universe it would be as easy to watch atoms and molecules interact as it is to watch our favorite actors on the silver screen, but that is not the case. We are constrained by fundamental physical limitations, such as the uncertainty principle, and where there are no fundamental limitations there are often practical engineering limitations.

Researchers have developed an impressive array of both quantitative and qualitative methods over the years in order to address this challenge. Experimental assays usually lead the charge in biological discovery because they help us find new molecules of interest and answer simple questions about them like where they are located in the cell and what they interact with. As the picture becomes more complex, so too do the methods we use and the kind of information they provide us, as do the fundamental physical challenges in achieving that picture.

There are many situations where it is important to have a more dynamic picture of these phenomena. Mutations in a protein, for example,

often lead to losses or increases in activity, either by affecting a protein's structure or its electrostatic interaction with a binding partner. Understanding the specific change to the structure or interaction can help us understand what is going on and design better therapeutics. Mutations to or overexpression of the epidermal growth factor receptor (EGFR), for example, are indicated in numerous cancers [1]. A single missense mutation in EGFR leads to the loss of effectiveness in one drug, Cetuximab, but not another, Panitumumab, that targets the same epitope [2]. Similarly, recent work in the study of G protein-coupled receptors (GPCRs) has shown that not all drugs are "created equal", meaning different drugs designed to target the same receptor and can have differing downstream effects in the cell depending on slight atomic changes in how they interact with the receptor [3, 4, 5]. The only way to understand the specific cause of such behaviors is to look at the atomic interactions. Information gleaned from such investigations can then be used to guide future drug design, leading to solutions that can maintain or recapitulate efficacy under such changes.
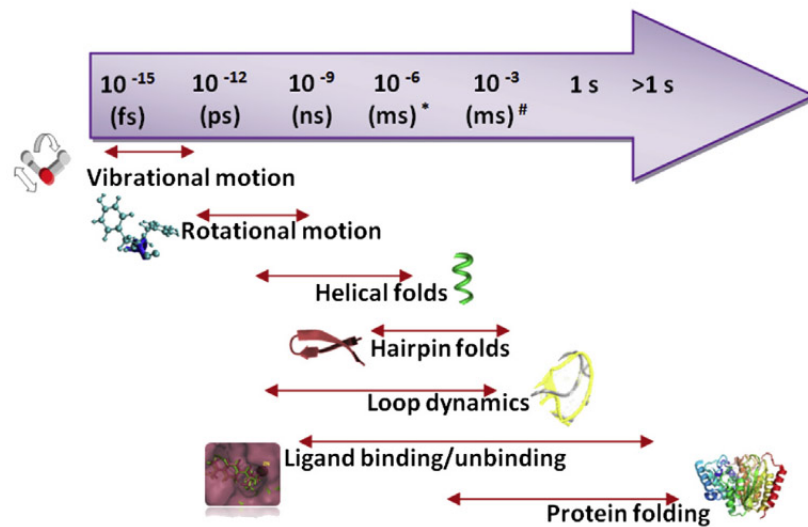

Figure 1.1: Rough timescales for various biological processes that involve proteins. Figure adapted from [6].

2

A critical part of this is the ability to characterize events that occur at atomic scale and that can span a huge range of timescales, from femtoseconds to hours or longer. Different kinds of protein motions span a large range of timescales (Figure 1.1) [7]. Basic protein motions range from picoseconds to milliseconds, proteins and small ligands interact on hundreds nanoseconds to seconds, and protein folding or protein-protein interactions can take substantially longer. A wide array of tools has been developed to characterize such motions, which we will cover in the next section.

1.1.1 Current methods in Structural Biophysics

A variety of experimental and computational techniques have been developed in an attempt to address these issues and build a complete picture of biomolecules at atomic or near-atomic resolution, each with their own advantages and limitations. They are generally limited in spatial or temporal resolution (Figure 1.2), and often come with other less obvious method-specific limitations.

The study of biology in 3D began in 1958, when the first crystal structure of a protein, Myoglobin, was solved [8]. Structures were added slowly over the decades, only reaching 507 structures by 1990, mostly small globular proteins. Since then the PDB has surpassed 100,000 structures [9], >88% of which are crystal structures [10].

Crystallography has provided an essential foundation for structural biology, but it does have limitations. Crystal structures are static pictures of dynamic structures, and represent either an ensemble average structure or the most stable conformation(s) susceptible to crystallization. Flexible loops are often missing from the spectra entirely [12, 13, 14, 15, 16]. Cryo-electron microscopy (CryoEM), a related technique useful for large, difficult to crystallize structures, also suffers from this. CryoEM also suffers from slightly lower spatial resolution than crystallography, but has been useful in determining several large macromolecular structures [17, 18, 19, 20].

3
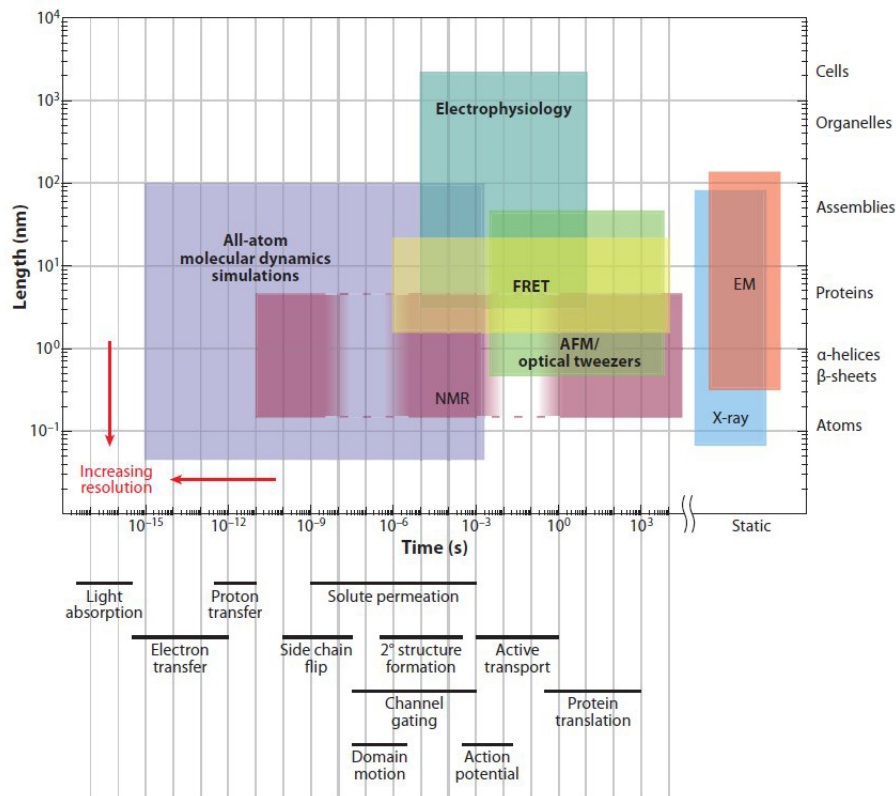
Figure 1.2: A plot showing showing the time and distance resolution of several common investigative methods used in structural biology. Below the x-axis are timescales of many common biological processes for comparison, and to the right of the distance scale are the sizes of molecular constituents and assemblies. Figure taken from [11].

The most successful experimental technique for investigating the dynamics of biomolecules, particularly proteins, has been nuclear magnetic resonance (NMR) spectroscopy, which has contributed approximately 10% of the structures in the PDB [10]. NMR can be used to determine the structure and dynamics of many biological molecules, including proteins, nucleic acids, carbohydrates, and many metabolites [21, 22]. A main ad-

vantage of NMR is that it can quantitatively describe populations and exchange rates between various conformers. Further, certain NMR methods can be used to find out binding sites for small molecules or other proteins, and so is useful for drug design [23, 24].

Much like crystallography, however, NMR also comes with notable limitations. First, it suffers from blind spots in the timescales of processes it can resolve. Carr-Purcell Meiboom-Gill Relaxation Dispersion (CPMG RD) and Rotating Frame Relaxation Dispersion (RF RD) are the most popular experiments used to investigate protein dynamics in the micro- to milli-second time scales by NMR. CPMG RD is well suited to detect exchange processes in the ca. 0.3-10 ms time window ($k_{ex} \approx 100 - 3000 s^{-1}$). RF RD can be used to study exchange events in the ca. 30-100 μs time window ($k_{ex} \approx 10,000 - 50,000 s^{-1}$). Therefore, processes in the (roughly) 100 ns to 40 μs and 10 ms to 100 ms ranges may be challenging to resolve (Figure 1.2) [25]. Further, molecules with a molecular weight larger than 35 kDa are progressively more difficult to resolve owing to overlapping of peaks and quicker magnetic relaxation [26, 27, 28]. This restricts the technique mainly to small soluble proteins, though techniques to escape some of these limitations have resulted in larger complexes being successfully analyzed [29, 30].

Several other techniques exist that also give important structural and dynamic information. Fluorescence resonance energy transfer (FRET) uses fluorescent probes that can give information on dynamics or structure as well. It can be used to study protein folding or the collapsed propensity of disordered proteins at sub-nanometer resolution [31, 32, 33], but because only two fluorophores are used the results are essentially limited to a one-dimensional projection of the process. Kinetics can be determined by using single-molecule techniques, and mutation scanning can help identify residues that play important roles in the process. Use of FRET for protein-small molecule binding is complicated by the effect the fluorophore may have on the small molecule, and is therefore better suited for protein-protein interactions, when possible. Small-Angle x-ray scattering (SAXS) can also give information on structure and dynamics [34], but like CryoEM it suffers from limited spatial resolution [35].

Other promising techniques are being pioneered, like the x-ray free electron laser (XFEL), which will potentially provide structure and dynamics at similar spatial resolution to x-ray crystallography. However, this will require highly specialized facilities such as the linear accelerator at Stanford, where one is being built.

In addition to these experimental tools, various computational tools have also been developed to both complement and corroborate their findings. Homology modelling is a technique that uses existing experimental structures in an attempt to build as-yet unknown structures. While this works fairly well when the basis structure and target are closely related, there are huge amounts of the protein space that have not yet been crystallized or cannot be crystallized, so the technique is of limited use in many cases [36]. Further, it cannot give dynamic information of any kind. Another computational technique known as docking similarly relies on existing structures. Docking attempts to determine binding sites of small molecules or other proteins. However, the inflexible nature of the structures severely limits the success to the simplest cases [37]. Flexible docking methods have improved the reliability and usefulness to some extent [38].

Other computational methods avoid the extensive use of experimental information by attempting to determine structures *de novo* via other means. Monte Carlo molecular modelling uses a sampling scheme whereby random, small changes are made to the system (such as dihedral angle changes in a folding study) and the energy of the new state is compared with previous states [39]. This can be quite successful for simple systems with a strongly funneled energy landscape, but can be prohibitively computationally expensive otherwise since the number of possible moves to be made grows exponentially with the number of degrees of freedom. The software program Rosetta is perhaps the most famous to successfully employ this technique [40, 41]. It has seen wide success in the de novo design of protein structures [42], the design of catalytic proteins [43], and self-assembling protein macromolecular structures [44].

It also might seem natural to use quantum mechanical simulations to investigate the fundamentally quantum processes of atomic interactions.

Quantum simulations are highly trusted in the research community and can view processes of arbitrary quantum complexity. However, they require extensive expertise to perform and can only simulate quantum processes up to nanosecond length, so are impractical for many studies. The size of systems that can be investigated is also severely limited, as scaling with number atoms is N^3 or worse [45]. A hybrid technique called QM/MM [46, 47], whose pioneers recently were awarded the Nobel prize for their work [48], is a much more practical approach that only treats a small subset of atoms in a simulation with quantum calculations. Quantum and QM/MM simulations are therefore best used when a specific, small set of atoms must be studied with high accuracy, such as when investigating the mechanism of a chemical reaction in an enzyme.

Molecular dynamics simulations, the prime investigative technique used in this thesis, are able to span a wide range of distance and timescales in principal. However, they were also not very practical until recently because they could not be performed on timescales long enough to see meaningful biological events like the folding of small proteins or the binding of ligands. This has changed drastically in the last five years, as outlined in the next sections.

1.2 High-throughput Molecular Dynamics and Markov State Modelling

1.2.1 MD

Molecular dynamics simulations represent atoms as point masses, and the interactions between them are determined by an empirically derived force field (FF) [49]. The force fields use a Newtonian representation of bonded and non-bonded interaction potentials, and the environment of each atom determines the forces upon it. The positions of the atoms at some time in the future are determined in a stepwise fashion based upon those forces, using common numerical integration schemes [50]. Various

schemes have been developed to allow for simulations in common thermodynamic ensembles, such as to maintain constant temperature (NVT), energy (NVE), or pressure (NPT), and are chosen depending on the requirements and goals of the simulations. The number of particles (N) is often fixed, though some schemes have been developed that add flexibility to this in order to maintain constant pH, for example [51, 52, 53].

The ability of MD simulations to accurately reproduce the physical world essentially comes down to the accuracy of the forcefields and the amount of simulation done, or sampling. There are several different common implementations of force fields which differ slightly in their formulas but for the most part have the same basic equation (Figure 1.3). The forces between atoms is controlled by bonded terms, which include

Figure 1.3: The basic equation for an MD force field. Adapted from Durrant and McCammon [54].

the inter-atom bonds, angles, and dihedral angles, and non-bonded terms, typically a van der Waals term and a Coulomb term. The most commonly used force fields for biology are Amber [55, 56] and CHARMM [57, 58]. Parameters for each term in the force field are derived empirically from ab initio (QM) simulations or are honed to match experimentally known parameters. For example, bond distances and angles are often known from x-ray crystal structures.

The force fields do differ in some important ways between these implementations, however. The constants that govern each interaction in the equations are different, owing to how each goes about parameterizing

8

those constants. The Amber force field parameterizes primarily from gas-phase QM simulations, which makes creating new parameters straight-forward but risks producing inaccurate parameters for in-water simula-tions [55, 59, 60]. The CHARMM force field, alternatively, parame-terizes using water molecules around the molecule, but the added com-plexity means properly creating new parameters requires experience [61]. Further, because performing QM calculations on every new molecule would be computationally prohibitive, both have also created generalized schemes involving atom types that map parameters onto new molecules based on similarity to ones already known. Amber has developed general AMBER force field (GAFF) [59, 62], while CHARMM has its general-ized force field (CGenFF) [61]. As a final difference, CHARMM also adds an additional term known as CMAP that adds nuance to the protein backbone terms [60, 63].

This simplification of inherently quantum processes may seem du-bious at first impression, and indeed the accuracy of the force fields was long an issue and led many to doubt results from MD simulations. Progress over the years has helped improve their accuracy and dispel these con-cerns significantly, particularly with respect to the parameterization for proteins, which are the main use of MD in biology. Several researchers have addressed issues in the protein backbone parameters in Amber force field, resulting in improved fitting to QM data in the Amber ff99SB force field [64], and improved helix-coil balance in ff99SB* and ff03* force fields [65]. Lindorff-Larsen *et al.* made further improvements to sev-eral side-chain parameters in Amber, resulting in ff99SB-ILDN [66]. The CHARMM force field had similar issues, resulting in over stabilized he-lices and salt bridges, which were corrected by Piana *et al.* resulting in CHARMM22* [67]. A comprehensive overview of these improvements up to 2012 is summarized by Lindorff-Larsen *et al.* [68]. A landmark study by the DE Shaw Research group showed that these forcefields can be used to fold a range of small proteins to <2 Åof their experimen-tal structures, providing strong evidence for the accuracy of these im-provements and the usefulness of MD for biological research at the same time [69]. Further improvements continue to be made. Most notable are

improvements that have been made to the interaction between water and the protein backbone, resulting in folded-unfolded ensembles that more accurately reflect experiments [70, 71].

In addition to the force fields, the other fundamental limitation of MD is that a small integration time step must be used, often around 5 femtoseconds or less. This is because the time step must be substantially smaller than the fastest motions in the system, namely bond vibrations, so that it reflects physical reality over the long term (known as convergence) [72]. However, the most basic protein motions like side chain flipping or loop motions take hundreds of nanoseconds or longer, meaning many orders of magnitude must be spanned in order to see simple motions in a single simulation [7, 73].

A standard consumer CPU has historically been unable to simulate even small systems (~25,000 atoms) on these timescales in any meaningfully useful amount of time. Fortunately, MD computations are highly parallelizable, and several approaches have been developed to take advantage of this. The traditional approach was to spread a simulation across multiple nodes of a cluster or supercomputer, but such resources are expensive and inaccessible for most users. With the introduction of generalized GPU architectures like CUDA and OpenCL and codes to take advantage of their built-in high parallelization [74, 75], performing a 1 μs simulation on a small system now takes just a few days at a cost that is easily accessible to most researchers.

1.2.2   HTMD: High-throughput MD simulations

While the ability to reach beyond 1 μs on a single GPU is important, many biological processes occur on tens of microseconds to milliseconds and beyond. Several enhanced sampling techniques have been developed to see such events, including metadynamics [76, 77, 78, 79], Accelerated MD [80, 81], or umbrella sampling [82, 83], among many others (for a nice overview, see [84]). However, they require aphysical biasing along a reaction coordinate or prior knowledge about a system, which in many

cases a researcher will not have. Coarse graining methods, which use highly simplified representation of biomolecules, have also seen extensive recent development [85, 86, 87]. However, this often results in a loss of accuracy (though impressive results can be achieved in certain cases, see [88, 89]). Specialized hardware, such as the Anton supercomputer [90, 91] or the MD-GRAPE [92], have been developed that can run single simulations on these long timescales, and these tools have been instrumental in helping to improve and verify the accuracy of force fields [66, 67]. However, those specialized computers are typically not economical or accessible for most researchers.

The most practical way to sample rare processes or those with long timescales is to use multiple parallel simulations, a method we refer to as high-throughput molecular dynamics (HTMD) [93, 73]. By starting multiple simulations from the same or different starting points, states and transitions between them can be extensively sampled such that meaningful statements can be made about their stability and the frequency. Running a single round or progressive rounds of parallel simulations increases the probability that these events can be seen and that adequate sampling can be achieved to see micro- to millisecond timescale processes.

Still, HTMD on its own is not always useful. The copious and disjointed nature of the data produced by HTMD studies means that making sense of it is a significant challenge. Further, it is often counterintuitive to newcomers that running multiple short simulations can allow you to investigate events that are much slower than the length of each individual simulation. This is possible thanks to Markov state models (MSM), which allow one to take advantage of the statistical probability of events.

1.2.3 Markov State Models

The most effective way to deal with the large amount of data generated by an HTMD study is by using Markov state model (MSM) analysis, which has seen a lot of development specifically for MD over the past few years. Markov state model theory for molecular dynamics is built

around prior work in transition networks [94, 95, 96, 97, 98]. The basic idea of MSM construction is to discretize your data into what are known as microstates via some metric, such as atom contacts, distances, or dihedral angles. Then, all of your data is binned by that discretization and transitions between them after a lag time are counted. The rate of these transitions gives information about the timescales upon which they occur. In the final stage, states that quickly interconvert are clustered together into macrostates, leaving only a few large, slowly interconverting states of interest (Figure 1.4). Transitions between these states are often processes like transitions from bulk to bound in the case of a protein-ligand binding, or from unfolded to folded in the case a protein folding study. From this model, important information like binding affinity or folding time can be determined and compared with experiments.

Multiple studies have successfully used MSMs to help reconstruct biophysical processes. They have primarily been used to investigate protein folding [94, 99, 100, 101, 102] and also protein-ligand binding [103, 104, 105]. In one study, for example, they were able to show that using just 50 μs of simulations they could correctly calculate the experimental binding affinity and kinetics [103]. However, there were several aspects of the methods in that work which made it difficult to generalize to other systems. The protein was restrained, and 2D and 3D spatial clustering of the ligand was used to build the MSM. Those who worked on using MSMs for protein folding encountered similar successes and limitations. While they could accurately approximate folding times, they found that RMSD based clustering was very limited in part because structures that were close in RMSD may not interconvert rapidly, hindering MSM construction [106].

Fortunately, a lot has been learned in the last few years about the proper construction of Markov state models, and new methods have been incorporated into the process. Clustering based on atom contacts or distances, for example, has proved to be much more effective than spatial clustering and removes the need for restraints on the protein such as those used in the Buch *et al.* work. For folding, using Cα-Cα distances and dihedrals is much more effective metric than RMSD.
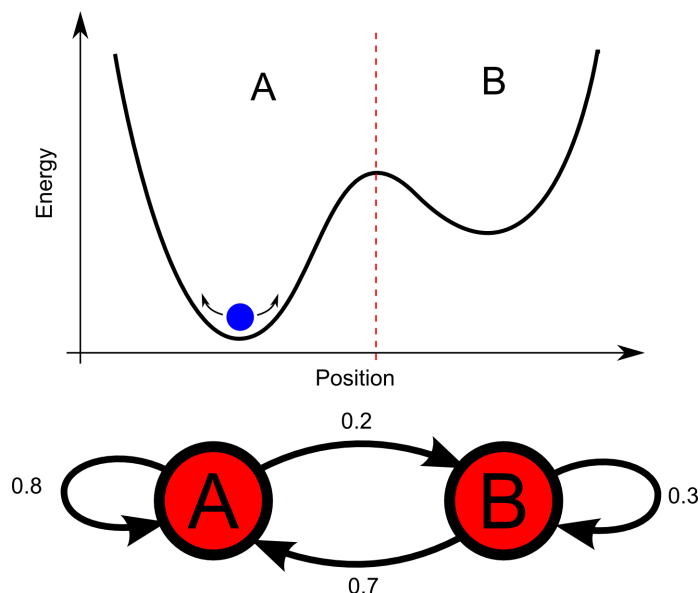
Figure 1.4: Basic overview of how an MSM works. Imagine a blue ball that is traversing an energy surface, such as the one pictured above, due to thermal fluctuations. We can split the energy surface at the saddle point, and then make note of the ball's position at some time intervals. If we count the transitions between the states after some lag time, we can build a count matrix, and thereby estimate the probability of transitions depending on the current position. The procedure is more complex in practice (we don't know where the saddle point is *a priori*, for example). One can imagine the above two-state model being an approximation of the transition from bulk (B) to bound (A) in binding, or unfolded (B) to folded in (A) in a folding.

Studies of folding and intrinsically disordered proteins provided additional insights into how to further improve MSMs. Clusters may be distant geometrically from the perspective of the clustering algorithm, but kinetically very close (fast interconverting). This would result in kinetically close states being clustered together when they were in fact dissimilar. A projection method known as time-sensitive Independent Component Analysis (tICA) was therefore incorporated into the process before

clustering in order to alleviate this problem [107, 108]. tICA projects the data along the system's slowest varying coordinates, which can then be fed to the clustering algorithms. This almost universally improves the accuracy of the Markov models.

Methods have also been investigated to use MSMs to help improve sampling. Adaptive MSMs have long been proposed as a way to ensure adequate sampling of processes seen and to explore as-yet unseen states [97]. Recent works have provided the first proof-of-principle in this direction, resulting in one case in a 10x decrease in the amount of simulation needed to properly characterize a binding process [109].

1.2.4 Molecular Recognition

A long-term goal has been to use HTMD to help better understand molecular recognition processes, which are the factors that result in specific interactions. This includes binding processes such as protein-protein interactions or the binding of small organic molecules to a target. Not only are the specific atomic contacts of the final bound complex interesting, but how molecules progress from unbound to bound, and whether that causes shifts in the shape of the protein or binding partner. At least three different models have been proposed for this: (i) the lock-and-key model, which proposes that ligands and the protein simply fit neatly into one another with no structural changes [110], (ii) the induced fit model, whereby the ligand induces a change in the protein towards some bound conformation [111], and (iii) the conformational selection model, in which the protein and binder sample a series of conformation until they fit together, and thus shift the ensemble of states toward that favorable conformer [112]. While this is a topic of intense debate and disagreement over the years [113, 114], there are copious examples of each type [115], and each system must be considered and studied on its own. The ability to better study how molecules progress from unbound to bound will provide a more concrete understanding of the factors that increase or decrease binding, and can lead to a better understanding of how certain diseases come about and how to treat them.

Currently MD simulations are the only way to a have a full atomic picture of such events, and so there is strong impetus to solidify the MD methods and produce proof-of-concept studies. There has already been substantial progress in this direction. As mentioned in previous sections, there is already evidence that MD works well for studying the binding of single molecules of interest to target proteins. However, questions have remained about how successful this will be when applied broadly. Our group has been able to makes strides in this direction, recently uncovering the bound poses from a 42 fragment screen against Factor Xa (unpublished) in which the primary binding modes overlapped well with available crystal structures in all but two of the fifteen cases, and the other two agreed with previous competition binding assays. This included affinity and kinetic data, as well as and showing intermediate weak poses.

In this doctorate I have focused on a different class of systems in which to lay the groundwork for studying molecular recognition, specifically membrane proteins and disordered proteins. These two general classes each have their own unique challenges and unknowns. In membrane systems, for example, the diffusion of the protein and lipids is exceedingly slow compared to molecules in water. For disordered proteins, the range of conformations that they sample is large, and the stability of those conformations can vary substantially. The only way to ensure that you have properly characterized such processes is with adequate sampling, which requires the kind of extensive simulation of HTMD. The following sections outline the background of these kinds of systems and the specifics about those we have studied.

1.3 Biological Systems Investigated

This section will give background on the kinds of systems studied in this doctorate using the techniques outlined above. A specific overview of each biological system is also given.

1.3.1 Membrane systems

Lipid membranes serve as fundamental barriers in cells, both between different cells and cellular compartments. They are therefore the site of numerous important processes and a wide array of biomolecules, with membrane proteins serving as the functional core of these. DNA and proteins have long received the lion's share of research attention, but we are beginning to learn that membranes and their lipids have direct roles in cellular processes. They are not homogenous, but heterogenous and the regulation of that can be quite complex [116]. Instead of being homogenous, lipids can partition into groups of differing composition, known as lipid rafts [117]. The composition of membranes is tightly controlled, and there can be significant asymmetry in the lipid composition of the leaflets [118]. Various lipids are principal actors in signalling [119], and are implicated in a number of disorders [120].

Still, lipids and membranes do not act alone, but are key players coordinated with other biomolecules for function, primarily membrane localized proteins. How lipids and membranes interact with and modulate membrane proteins is increasingly being studied, but there are clear indications that they must be considered together [121]. Membrane proteins make up 20-30% of the proteins in a cell [122]. Roughly 60% of all approved drugs target membrane proteins, with more than half of those targeting a specific class of receptors known as G protein-coupled receptors (GPCRs) [123]. They are frequently targeted because they are outside the cell and can have dynamic downstream signalling effects, thus serving as a convenient point of regulation. They are therefore of great interest from a therapeutic perspective, and any methods that increase our ability to understand them are welcome.

Membrane proteins may be easy to target with drugs, but they pose a significant challenge to study from a structural standpoint compared to proteins that are stable in water. The fact that their preferred state is in a membrane means that they must be studied in this context, or in a context that mimics membranes and does not disrupt their structure or

function. This makes them more challenging to crystallize for x-ray analysis. Of the membrane protein crystal structures that do exist, most are from bacteria, and unlike many water soluble proteins, using a strategy of heterologous expression in bacteria is often unsuccessful [124]. Some, such as some GPCRs, are very lowly expressed even in native tissues, adding to the complications. Other methods to study structure, such as NMR, have difficulty due to the membranes influence on the spectra, but can provide important insights [26, 125, 24, 126]. Fortunately, the number and diversity of membrane protein structures determined is steadily increasing [127].

Studying the binding of ligands to membrane proteins with many experimental techniques is also a challenge. Due to the difficulty in getting a structure in the first place, it may not always be practical to use crystallography to find a binding pose. Various NMR methods, such as solid-state or saturation transfer difference (STD) methods, can be used to find binding poses or other locations of protein-ligand interaction in membrane proteins [128, 129, 130], or to look at their conformational fluctuations [131]. Still, these often provide information only on the most stable state, and not intermediates. Methods like radioligand binding, isothermal titration calorimetry (ITC) and SPR can be used to learn affinity or kinetic properties of binding [132], but naturally provide only bulk information and no structural detail. These methods must be used in conjunction to build a picture of what is going on, as they suffer from individual time or spatial resolution limitations, or are simply substantially more difficult to carry out on membrane proteins.

As structures become available, simulations can potentially resolve some of the issues faced by these experimental techniques, and have already been used extensively to understand membrane proteins in ways that would be difficult or impossible without them. MD has been used to study the binding of several drugs to the β_1- and β_2-adrenergic receptors [133], and the allosteric modulation of M2 receptors has also been studied [134]. It has also been used to understand the dynamic activation of M3 and β-adrenergic GPCRs [135, 136], which is something that is becoming increasingly important for the design of better medicines.

17

They have also been used to uncover the conformational changes important to EGFR activation [137, 138, 139]. Finally, they have been used to uncover the mechanisms of ion channel function [140, 141, 142]. All of these studies would be difficult or impossible without MD, and are excellent examples of how simulations can make key insights into membrane protein function.

The possibility of getting full atomic detail of binding is intriguing, and is what drove the following studies. In both, we used MD to study the binding of lipid molecules to membrane proteins in order to push the boundaries of what has been done and to better understand these processes. The goal was to understand how these molecular recognition processes occur and how they are similar or different from ones previously studied.

FAAH: Fatty Acid Amide Hydrolase

The endocannabinoids (ECs) are among the most abundant neurotransmitters in the brain, and play a role in a wide array of physiological and medically important processes, such as pain and inflammation, energy metabolism, neurological modulation and even cardiovascular function [143]. They are present in numerous types of neurons, and help modulate ion channels and neurotransmitter release [144]. It has therefore been proposed that successful strategies to regulate the EC system could have a range of therapeutic applications [145, 146, 147]. However, this broad range of functions has meant that drugs targeting the EC system can have undesired side effects, such as in the case of Rimonabant, which successfully treated obesity but resulted in unacceptable levels of anxiety and depression [148]. Any work that provides a better understanding of the system may help overcome these limitations.

The main signalling molecules of the EC system are the lipids anandamide (AEA) and 2-arachidonoylglycerol (2-AG), which are degraded by Fatty acid amide hydrolase (FAAH) [149] and monoacylglycerol lipase (MAGL) [150], respectively. Both FAAH and MAGL are monotopic integral membrane proteins, and are located in different sides of the synapse

(Figure 1.5), with FAAH being located primarily postsynaptically to CB1 receptors [151]. FAAH is a homodimer and a member of the serine hydrolase family of enzymes, and cleaves anandamide into ethanolamine and arachidonic acid. As anandamide is a lipid, its binding to FAAH proceeds from the membrane. Since it terminates EC signalling, its inhibition is a potential pathway to broad modulation of the EC system.

Motivating our work with FAAH, experimental collaborators had indications that the amount of cholesterol present in the membrane modulates the enzymatic activity of FAAH. Increasing the amount of cholesterol in the membrane increases the activity of the enzyme. This seemed peculiar, as cholesterol rigidizes membranes and should decrease diffusion. It was therefore unclear by what mechanism it increases the activity of FAAH. Did cholesterol interact directly with FAAH, or did the presence of cholesterol change the membrane in some way? Did it change the positioning of the enzyme somehow?

This was a good opportunity not only to help them clear up confusion regarding FAAH itself, but also to test whether MD can be successfully used to simulate the binding of lipids to membrane proteins. While numerous examples exist of MD being used to simulate ligands that diffuse freely in water, there was still no case of lipid binding. Ligands diffuse much faster in bulk water than lipids do in membranes, and therefore even in a single long simulation may not see even a partial binding event. For lipids, the timescale is much slower, and the only way to see binding events is via newer methods like HTMD. Therefore, we were in a unique position to perform the study of FAAH.

S1P$_1$R: Sphingosine-1-phosphate receptor

Members of the G protein-coupled receptor superfamily serve to regulate a vast array of functions in the human body, from neurotransmission to differentiation to adhesion, and are the target of over 30% of all approved drugs [123]. They have therefore been the focus of intense research since
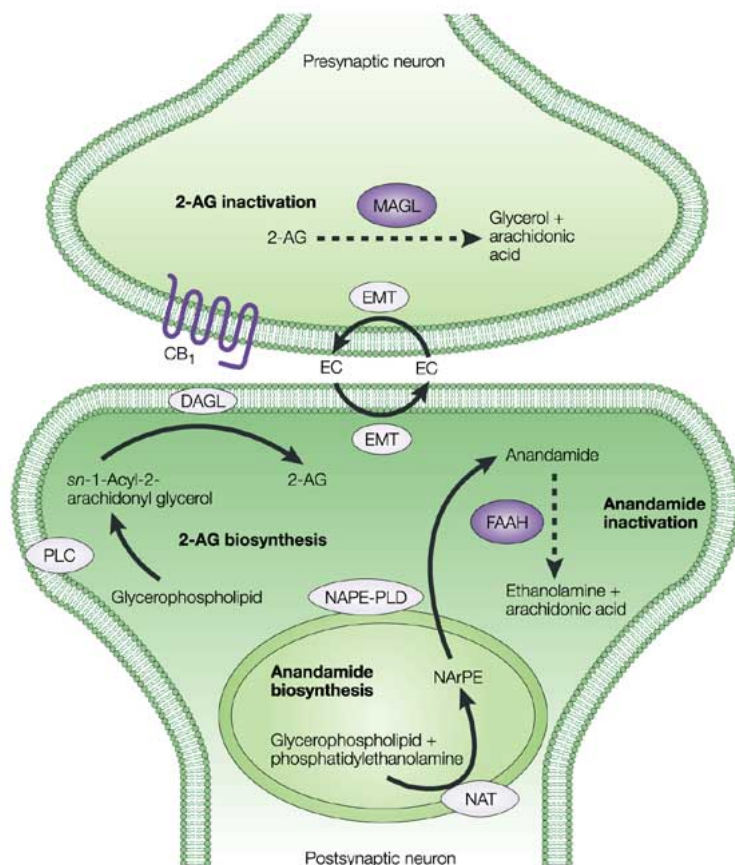
Figure 1.5: Overview of the endocannabinoid system. Adapted from [145]

their discovery. However, their structures have historically been difficult to study because they are lowly expressed and were hard to crystallize [124]. Recent advancements in methods have led to an explosion of structures becoming available in the past five years [152]. There are now

over 75 structures available that span almost all branches of the family.

One of the more interesting structures crystallized was that of the Sphingosine-1-phosphate receptor 1 (S1P$_1$R), which was the first lipid GPCR to be crystallized, in 2012 [153] (Figure 1.6). The S1PR family is important in endothelial cell cytoskeleton structure, maturation, and vascular tone and a host of other cell regulatory functions [154, 155, 156]. It also has a role in the maturation and migration of lymphocites in immune response [157]. Activation of S1P$_1$R with the drug FTY720 affects vascular permeability, and can be important to the delivery of chemotherapeutic agents to the brain or the treatment of multiple sclerosis [158, 159, 160]. Alternatively, S1P$_1$R has a role in promoting cancer cell growth, motility, and angioneogenesis, and its inhibition could therefore be a potential treatment [161, 162].
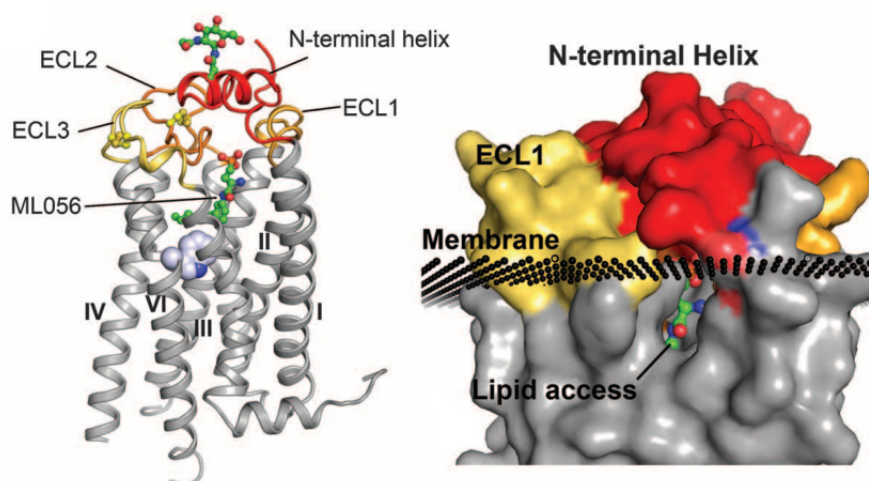

Figure 1.6: Overview of the S1P$_1$R receptor structure (left) and proposed binding port (right). Adapted from [153].

While crystal structures give a wealth of essential insights into receptors, they provide only a static picture of something that is involved in

many dynamic processes. Recent studies have shown that the activation and inhibition of the GPCRs is not simply a digital on/off process, but is rather nuanced and complex dynamic process in which an activator or inhibitor is a key player [135, 163]. Understanding how a ligand binds to and modifies a receptor, therefore, is important for basic biology as well as for the development of therapeutics.

While it is easy to determine affinity and kinetics for many water-soluble ligands and proteins, it is substantially more difficult for membrane proteins and the GPCRs in particular. Numerous NMR studies have been undertaken to determine structure [164, 165], dynamics [131], and ligand binding in GPCRs [129, 128]. Still, such methods often cannot describe the full binding process, from intermediates to the bound pose, and look at only small subset of atoms. Molecular dynamics (MD) simulations allow for the study of the entire protein and ligand interaction at atomic scale, and have recently been used to visualize such processes with impressive results [103], including with GPCRs [133], though we are the first to do so with a lipid ligand.

To better understand the binding process of ligands to this GPCR, and to show that it was indeed possible to undertake such a study, we chose to study the binding of the ligand ML056 to the receptor. The structure of ML056 (a.k.a. W146) is similar to the endogenous ligand S1P, and therefore lessons we would learn from it would likely be applicable to the binding mechanism of the native ligand as well. Much like our work on FAAH, the motivation for working on $S1P_1R$ went beyond providing insights purely focused at the system itself.

1.3.2 Disordered proteins

Note: Parts of this section are taken or adapted from [166] and Publication 3.3.

Another area where HTMD simulations can provide invaluable insights is with intrinsically disordered proteins. Intrinsically disordered proteins (IDPs) are proteins that lack or have highly-transient secondary

and tertiary structure. This makes them particularly difficult to character-ize by traditional biophysical techniques like x-ray crystallography, and they were ignored for many years because of this. This changed around the turn of the century as their importance was acknowledged thanks to a few seminal works [167, 168, 169, 170]. The first clues to their exis-tence came from crystal structures with missing sections in their electron density maps, in some cases parts critical to function [12, 13, 14, 15, 16]. This, along with data from NMR and CD experiments, led to the creation of a database for disordered regions like DisProt [171] and inspired tools to try to predict disorder from sequence alone (PONDR) [172]. There are many such predictors now, and even meta predictors like PONDR-FIT that combine them [173, 174, 175].

Extensive investigation since then has made it clear that IDPs have frequent and important roles in biological processes. Disorder is found in both prokaryotes and eukaryotes, but is higher in eukaryotes, where disordered regions are found in more than 50% of proteins [176]. Dis-ordered regions are enriched in regulatory and signaling proteins, and are less commonly found in proteins responsible for metabolism, biosyn-thesis or transport. Disordered regions are also frequent targets of post-translational modification, and post-translational modifications in proteins participate in many fundamental cellular processes [177, 178, 179]. They affect at least one-third of all eukaryotic proteins [180, 181], preferen-tially targeting intrinsically disordered protein domains [180, 181].

Having such prevalence in key regulatory functions in the cell means they are commonly found to play roles in various diseases [182]. They are found mutated in numerous cancers [183], unexpectedly common in cardiovascular disease [184], and they are common components in the fibrils of various amyloidoses like Alzheimer's disease, Parkinson's dis-ease, and diabetes [185, 186]. There has therefore been great interest in studying their properties so we can better understanding of how and why they cause or participate in such diseases.

When it comes to binding or molecular recognition, disordered pro-teins appear to have various functional differences from folded proteins

23

or small molecules. Among them is the ability to bind to multiple different binding partners in varied conformations, and to form weak but highly-specific interactions [187, 188, 189]. Their propensity to undergo post-translational modification is likely integral to this. There is also significant debate about the implications of disorder for the thermodynamics of binding, and whether disordered proteins undergo conformational selection, induced fit, or some complex combination of both [190, 191, 192, 193]. It has also been proposed that their disorder-to-order transitions may result in novel allosteric mechanisms [194].

Despite all this progress in understanding them, IDPs are still difficult to study from a biophysics point of view. The methods used suffer either in limitations in their scale or time resolution. As already mentioned, crystallography can give accurate information on atomic positions, but is limited by the fact that positions of atoms must be stable, or at least transition only slowly between a few positions. Nuclear magnetic resonance (NMR) methods can give general information about residual secondary structure or transiently formed long range contacts [195], as well as the time scale of conformational transitions. However, the information is ensemble averaged and, despite recent advances [25, 22], limitations on the accessible time scale remain. Other methods such as SAXS and single-molecular FRET can help understand the degree of collapse. In short, experimental methods have clear limitations in their ability to give detailed information about the states and transitions of IDPs.

The challenges faced by the methods above stress the need for new approaches. Secondary structural motifs like α-helices and β-hairpins form on the 0.1-10 μs timescale, and even the fastest folding proteins take multiple microseconds to milliseconds to fold [69, 196]. Meaningful transitions in IDPs will likely occur on similar timescales, so any technique that is to fill this void must be able to identify transitions and metastable states formed on these timescales or longer. Long timescale, explicit solvent MD simulations is perhaps just the tool to help understand IDPs. Micro- to millisecond MD simulations have in recent years significantly contributed to our understanding of dynamic protein conformations [69, 197, 198, 199], including for some disordered proteins [200, 201]. We

believe that HTMD is just the tool for this, and work detailed below are our first successes in this direction.

KID: Kinase Inducible Domain

The kinase inducible domain (KID) is one of the first disordered proteins to have been studied, and also one of the most extensively studied. KID is a 60 amino acid domain of the CREB transcription factor, and has numerous binding partners, most famously the KIX domain of CBP (Figure 1.7) [202].

KID is known to be disordered in solution and to form two alpha helices upon binding to KIX [203], a process that involves at least one binding intermediate [204]. Binding of KID is regulated via phosphorylation of S133 in the αB helix, which increases its binding affinity 40-fold (binding of residues 119-147 to KIX). Interestingly, phosphorylation barely affects the fraction of folded αA and αB helices in solution [205]. Computational studies on KID protein using replica exchanged implicit solvent MD simulations and short all-atom MD simulations further showed that KID is largely unstructured and phosphorylation barely affects its helical propensity [206, 207]. Moreover, various computational studies using coarse grained models, short high temperature simulations, and Gō models suggested that binding of KID to KIX initiates at the αB helix [207, 208, 209, 210]. Mutation studies suggested interactions between the phosphorylated serine and residues on KIX are the main driving force of this increased affinity [203]. However, the mutation of S133 to a negatively charged residue such as glutamate (often considered to mimic interactions with amide NH, lysine and arginine residues [211]) cannot recapitulate pKID activity even marginally [212].

Post-translational modifications (PTM) such as this are highly common in disordered regions like KID, and it is unclear what effect they have on their conformations or what implications that may have for binding. Histone tails are known to be extensively modified, and the amount of modification can result in the binding or release of the DNA it is bound to [213]. Phosphorylation is a very common PTM and is known to have
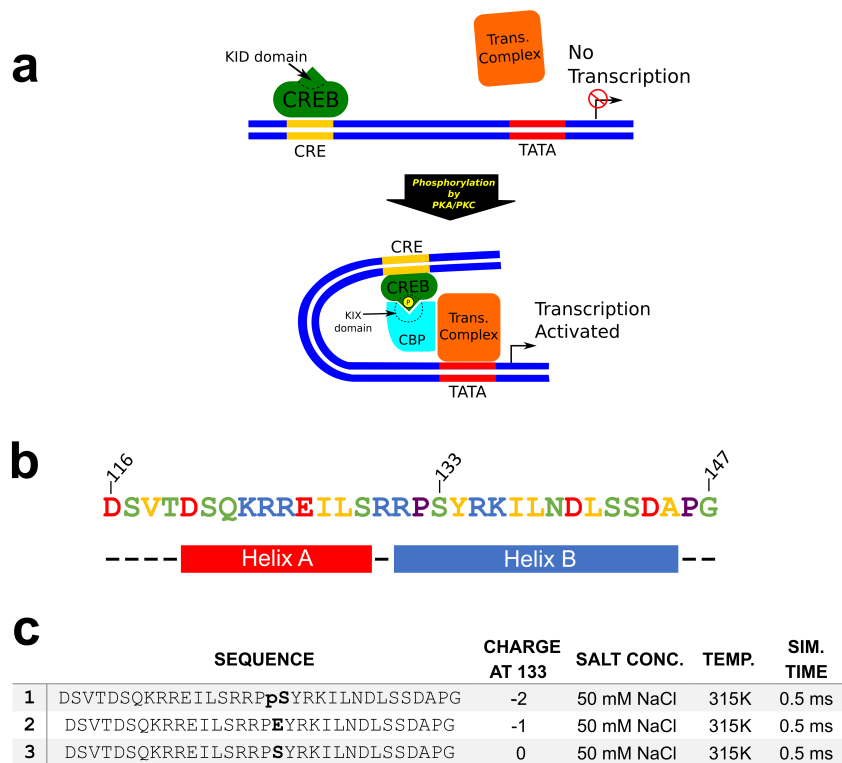
Figure 1.7: Overview of the activation of KID and its binding to the KIX domain of CBP (a), along with a short sequence of KID known to bind to KIX (b).

numerous effects on proteins. It can induce conformational changes [214], promote order-disorder transitions [215], and modulate binding via electrostatic interactions with partners [216]. However, something that has been significantly less well studied is how phosphorylation can regulate the conformational kinetics of proteins, and what effect this may have on interactions with binding partners. This is especially interesting for protein domains that can transition between unfolded and folded, but also difficult to address because it means that lowly populated, transient states

must be adequately characterized.

In an attempt to understand how phosphorylation modulates disordered states of proteins and their binding, we used the KID system as a model and determined the conformational kinetics and energetics of the domain before and after phosphorylation at atomic resolution. We chose to study an experimentally well-characterized disordered fragment of the KID domain of transcription factor CREB [202] (residues 116-147). We used HTMD simulations to perform over 1.7 milliseconds of aggregated simulation time of the phosphorylated, non-phosphorylated and S133E mutant forms of KID. Our initial goal was simply to ensure that we could reproduce experimental observables. However, it became apparent quickly in the course of our research that we were observing processes that had yet to be observed, and as we make clear in publication 3.2 of this thesis, could have broad implications for the binding of IDPs.

Chapter 2

OBJECTIVES

The main objective of this doctorate has been to pioneer the use of high-throughput molecular dynamics (HTMD) simulations to study the behavior of complex biological systems and binding processes they are involved in. Molecular dynamics is uniquely poised to study systems where current experimental techniques are limited and traditional simulation methods are inadequate. This is true particularly for membrane systems and proteins that are disordered or transiently ordered, and these were therefore chosen as the focus of this doctorate. We outline these objectives below:

2.1 Establish foundations and feasibility of using simulations to study binding in membrane systems

Proteins and other biomolecules localized on membranes are particularly difficult to study at atomic scale by experimental methods, and is a potential role that MD can fill. However, there are many challenges for MD that must be understood and overcome. The motions of molecules

in and around membranes, such as the diffusion of lipids in the membrane, is drastically reduced when compared with diffusion in water. Further, with the increased number and variety of interacting molecules, there are uncertainties regarding the ability of force fields to accurately reflect their interactions. Finally, the ability to use MD for the binding of lipid molecules was unknown, but important for the study and development of therapeutics.

High-throughput MD is an ideal method to address these questions. We chose to tackle this by finding the most challenging problems, which was to study the binding of lipid ligands to two different membrane proteins, FAAH and S1P$_1$R. FAAH is a monotopic membrane protein that terminates the endocannabinoid system by hydrolizing the endogenous ligand anandamide. A key challenge was reproducing the binding of this ligand, and also to explain the mechanism by which cholesterol modulates this binding. In another work, we studied the binding of an lipid inhibitor to the GPCR S1P$_1$R.

2.2 Investigate behavior of disordered proteins on biologically meaningful timescales

Disordered proteins or protein domains that are only transiently ordered are another area where HTMD can serve a vital role for biological research. There are few good experimental techniques that can characterize their motions and metastable states, particularly motions that occur on certain timescales. MD is only limited by the upper limit that it can reach, and with the methods outlined in section 1 we are able to make some the first significant simulations in this area.

We have used exhaustive MD simulations of disordered proteins to better understand their behavior. Publication 3.2 shows that these simulations have uncovered new, previously unseen processes and led to the proposal of a new mechanism by which post-translational modification my lead to modulated binding. Publication 3.3 is a review in which we discuss the significance of this work in the context of other such works.

Chapter 3

PUBLICATIONS

3.1 Membrane Lipids Are Key Modulators of the Endocannabinoid Hydrolase FAAH

Enrico Dainese, Gianni De Fabritiis, Annalaura Sabatucci, Sergio Oddi, Clotilde Beatrice Angelucci, Chiara Di Pancrazio Toni Giorgino, Nathaniel Stanley, Michele Del Carlo, Benjamin F. Cravatt and Mauro Maccarrone
Biochemical Journal 457, no. 3 (2014): 463-72.

Summary

In this project we worked with experimental collaborators to better understand how a membrane protein, Fatty Acid Amide Hydrolase (FAAH), is modulated by the membrane itself. Our collaborators in the Dainese group had seen that increasing the amount of cholesterol the membrane would increase the activity of the enzyme. This was peculiar, and there were many arguments for or against why this might make sense. Cholesterol rigidizes membranes, which we thought might increase the stability of the enzyme or change its resting position in the membrane. But cholesterol also decreases diffusion, which could have actually decreased the rate of catalysis. An alternative possibility was that cholesterol has some

direct interaction with the enzyme. Indeed, our simulations confirm this later hypothesis, and we showed that cholesterol preferentially interacts with the enzyme where the endogenous ligand enters. Further, multiple binding events in our simulations show that cholesterol interacts with the ligand and a salt bridge on the enzyme during binding.

Dainese E, De Fabritiis G, Sabatucci A, Oddi S, Angelucci CB, Di Pancrazio C, Giorgino T, Stanley N, Del Carlo M, Cravatt BF, Maccarrone M. Membrane lipids are key modulators of the endocannabinoid-hydrolase FAAH. Biochem J. 2014 Feb 1;457(3):463-72. doi: 10.1042/BJ20130960.

3.2 Kinetic modulation of a disordered protein domain by phosphorylation

Summary

Intrinsically disordered proteins have recently been acknowledged to have important roles in many cellular processes, but because they have no stable structure and can change conformations on a wide variety of timescales they are difficult to study with many experimental techniques. In this work we give one of the first examples of how MD can be used to fill these gaps left by other methods by uncovering a long-lived metastable state in the KID protein that had never been seen before. This state arises due to a post-translational phosphorylation, and cannot be reconstituted by mutating to glutamate, indicating that the phosphate is specifically required for this state to arise. We further show via a kinetic model that such a kinetic change can have real consequences for binding, resulting in a 10-fold increase in binding even in cases where there is no substantial change in populations of different states.

Stanley N, Esteban-Martín S, De Fabritiis G. Kinetic modulation of a disordered protein domain by phosphorylation. Nat Commun. 2014 Oct 28;5:5272. doi: 10.1038/ncomms6272

3.3 Progress in studying intrinsically disordered proteins with atomistic simulations

Nathaniel Stanley, Santiago Esteban-Martín, Gianni De Fabritiis. *Progress in Biophysics and Molecular Biology* Submitted, under review.

Summary

In this review we discuss the history of the study of disordered proteins and recent results in atomistic simulations that have provided new insights into their behavior. Simulations have uncovered behaviors that were previously unseen or invisible to other methods, and have included several discoveries that have therapeutic importance. We focus primarily on simulations that exceed tens of microseconds of sampling, including works with several milliseconds worth of sampling.

Stanley N, Esteban-Martín S, De Fabritiis G. Progress in studying intrinsically disordered proteins with atomistic simulations. Prog Biophys Mol Biol. 2015 Oct;119(1):47-52. doi: 10.1016/j.pbiomolbio.2015.03.003

3.4 High throughput molecular dynamics for drug discovery

Summary

Molecular dynamics is progressively becoming a tool that will be useful to understand biological disorders and to design therapeutics. Recent advancements in GPU hardware, simulation software, and commodity cloud services mean that running extensive, parallel simulations, or high-throughput MD (HTMD), is now practical for almost any researcher. Further, advancements in Markov state model analysis make analyzing HTMD data much more tractable. In this review we cover all these key advancements and discuss recent works that show that HTMD can now be used for drug discovery purposes, such as to run fragment screens, assess the binding of lead compounds, or to better understand the behavior or biological systems.

Stanley N, De Fabritiis G. High throughput molecular dynamics for drug discovery.
In Silico Pharmacol. 2015 Feb 13;3:3. doi: 10.1186/s40203-015-0007-0.

3.5 The pathway of ligand entry from the membrane bilayer to a lipid G protein-coupled receptor.

Nathaniel Stanley, Leonardo Pardo, Gianni De Fabritiis.

Summary

Membrane proteins make up 30% of the proteins in the cell, but localization in the membrane makes studying them challenging in many cases. This is particularly true when studying the binding of ligands, and even more so when the ligand is itself a lipid. We have undertaken the first study of the binding of a lipid ligand to a G protein-coupled receptor in order to verify its proposed binding pathway and uncover important interactions along it. We show that the lipid inhibitor ML056 binds via a port at the membrane surface, and that this binding is regulated by residues at the top of transmembrane helix 7 and residue in a flexible N-terminal helix.

Stanley N, Pardo L, Fabritiis GD. The pathway of ligand entry from the membrane bilayer to a lipid G protein-coupled receptor. Sci Rep. 2016 Mar 4;6:22639. doi: 10.1038/srep22639

Chapter 4

DISCUSSION

The compendium of work highlighted in the previous chapter demonstrates that we have been able to use HTMD in several cases to elucidate complex biological processes that would be difficult to see by other means. With HTMD we have been able to get a picture of molecular recognition processes at atomic scale and gain broader insights into how such processes occur. Here we discuss the implications of these results, ongoing work in these areas, and future challenges.

Establish foundations and feasibility of using simulations to study binding in membrane systems

We studied lipid binding to two different membrane proteins, FAAH and $S1P_1R$. In the FAAH system we not only achieved our primary objective of reproducing a binding event, but we saw multiple partial binding events. We were also able to propose a mechanism by which cholesterol modulates the enzyme by direct interaction. A 2D histogram from all the simulations showed that cholesterol preferentially interacted with the enzyme near the ligand entrance port. While diffusion of the lipids was indeed slow, the HTMD method increases the probability that rare and slow events could be fully seen, and this work exemplifies that. Further,

the fact that the simulations were unbiased indicates that the force field was accurate enough to result in these events being seen.

Our simulations of $S1P_1R$ were similarly successful. We were able to spontaneously reproduce the crystal bound pose of the lipid ligand ML056 to the receptor while showing important stages along the binding pathway (Figure 4.1). We further showed that instability in the N-terminal helix may be an important factor in the binding process. The ligand bound in the study, ML056, shares many structural similarities to the endogenous ligand, S1P, and therefore the lessons we have learned are transferable to it as well. In particular, residues R292, E294

Figure 4.1: The main binding pathway for ML056 to $S1P_1R$.

and Y295 at the top of helix 7 likely serve as a gating mechanism to the binding site that specifically recognizes the head group of the lipid. The fact that mutation of these residues substantially reduces activation of the receptor supports this [217]. Loss of those residues likely results in either an inability of the receptor to properly recognize and capture S1P ligand, or an inability to keep out non-productive binders (molecules that "clog" the receptor, but don't activate it).

Still, while these membrane simulations were a success from a qualitative standpoint, bigger questions and challenges remain. The accuracy of the simulations will be fully supported once the affinity and kinetics of such processes can be estimated and compared with experiment. However, development remains on both sides in order for that to happen. In the $S1P_1R$ system we were able to build an MSM and estimate binding affinity, but we lacked experimental measurements to compare to. Such

measurements are rare and difficult to obtain, and even more difficult to interpret.

Investigate the behavior of disordered proteins on biologically meaningful timescales

Our work on KID is one of the first studies of disordered proteins using unbiased MD on timescales longer than ten microseconds, and the most extensive to date with over 1.7 milliseconds of aggregate simulation. We found that phosphorylation at residue S133, a known phosphorylation site, results in a 60x slowdown in conformational exchange in KID, and that this is not recreated in the S133E mutant (Figure 4.2). This phenomena had remained hidden from other methods such as NMR, and is a clear example of how HTMD methods can provide key insights to biological processes.

Figure 4.2: A schematic representation of kinetic modulation of KID by phosphorylation.

Our observation of the kinetic slowdown made us wonder whether such a process could have consequences for binding, and led us to propose a kinetic model that shows that such change could alone result in

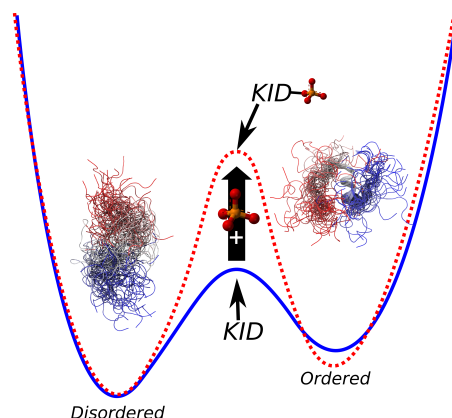increased binding affinity to a known binding partner (Figure 4.3). This is the first time such a model has been proposed, and while its applicability to KID/KIX binding remains to be fully fleshed out, it is joined by other recent discoveries that support the idea that factors affecting disorder can have fundamental influence on binding [218]. Simulations have been used by others to show that post-translational modification such as phosphorylation or mutations can result in changes in disorder propensity that are in fact important to binding [137, 219].

Whether the kinetic effect we have uncovered in this case is a common mode of modulation remains an open question, though it is likely simulations will be needed to fully understand this phenomenon. It is also unclear if it is important for the modification to be charged. Other post-translational modifications such as methylation of lysines result in changes in charge of the residue, and so even if charge is integral part of the mechanism it would not be unique to phosphorylation sites.

The question remains whether HTMD is ready to be used to visualize the binding of IDPs to their target proteins. In the case of KID/KIX, the on rate is roughly $10^6 M^{-1} s^{-1}$, meaning that it would likely require several milliseconds of simulation to see a binding event. This is at the limit of what is currently possible, and the fact that it is known to have

Figure 4.3: A kinetic model we proposed in [166] that shows that a 100x slowdown between productive and non-productive binders can result in a 10x increase in binding affinity.

more complex kinetic binding profile [204] could further complicate such a study. Other cases are substantially more tractable, such as in the case of c-Myb/KIX. The k_{on} rates for c-Myb binding to KIX are faster than some of the fastest folded protein-protein pairs or small molecules, on the order of $10^7 M^{-1} s^{-1}$ [220]. We have already begun some work in this direction with positive results. As adaptive MSM methods become more sophisticated, this kind of study should become tractable for even the more complex cases.

A range of other questions remain as well. Is the three-state binding (or more) such as that seen by KID/KIX normal for IDPs? It is also unclear by what mechanism IDPs bind to their interaction partners. Do IDPs typically follow a conformational selection model (where they bind only after reaching a required conformation), or whether they begin to bind and then slowly fold into their bound conformer. Work done on KID/KIX already suggests that for IDPs the situation may be more complicated, and perhaps progresses through an amalgam of both mechanisms [192]. The αB helix appears to bind first, and it is possible that it must be folded in order to initiate a productive binder [207, 208, 209, 210]. However, thereafter it may then progress more like an induced fit mechanism, whereby the αA helix and other parts slowly fold and form their native contacts. Is this how it happens, and is that a common mechanism for IDP binding? Atomistic simulations such as those outline in this thesis will help us answer such questions.
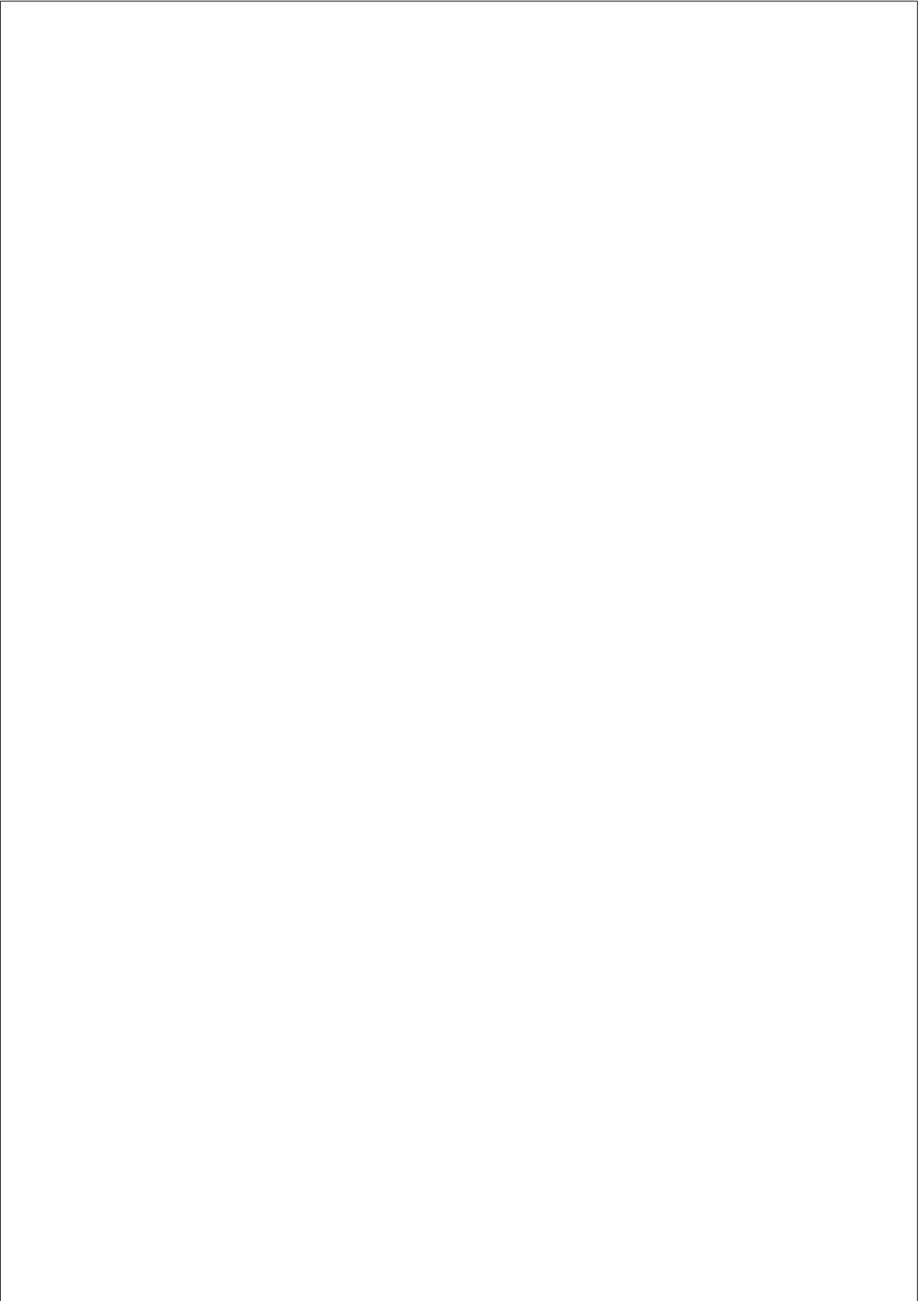
Chapter 5

CONCLUSIONS

1. High-throughput molecular dynamics simulations are a useful tool for studying the slow processes seen in many membrane system. It is possible to investigate the binding of lipid ligands such as anandamide or ML056 and uncover important events along their binding pathways.

2. Anandamide binds spontaneously to fatty acid amide hydrolase in simulations. Cholesterol directly interacts with FAAH near it's binding port to facilitate binding.

3. The lipid inhibitor ML056 binds to the $S1P_1R$ receptor via a multi-stage process involving key recognition residues and a flexible n-terminal helix that caps the binding site.

4. High-throughput MD simulations combined with new methods in Markov state modelling analysis has uncovered an as yet unseen behavior in a well studied disordered domain, the kinase inducible domain.

5. A novel mechanism uncovered in KID protein led to the discovery of a model under which purely kinetic changes can lead to increased binding affinity. This novel mechanism may be broadly applicable to a wide range of systems, particularly where conversions

betweens various different states has consequences for binding or other activities like catalysis.

Chapter 6

LIST OF COMMUNICATIONS

Articles

1. Dainese, E. et al. Membrane lipids are key modulators of the endocannabinoid-hydrolase FAAH. Biochemical Journal 457, 463-472 (2014).

2. Stanley, N., Esteban-Martín, S., De Fabritiis, G., Kinetic modulation of a disordered protein domain by phosphorylation. Nature Communications, 5 (2014) DOI:10.1038/ncomms6272

3. Stanley, N., Esteban-Martín, S., De Fabritiis, G., Progress in studying intrinsically Disordered proteins with atomistic simulations. Progress in Biophysics and Molecular Biology. Submitted

4. Stanley, N., De Fabritiis, G., High throughput molecular dynamics for drug discovery. In Silico Pharmacology. 3:3, (2015) DOI:10.1186/s40203-015-0007-0

5. Stanley, N., Pardo, L., De Fabritiis, G., The pathway of ligand entry from the membrane bilayer to a lipid G protein-coupled receptor.

Talks

1. Lipid Binding through Membranes Using Molecular Dynamics, International Workshop on Computational and Theoretical Modeling of Biomolecular Interactions June 5th 2013, Moscow/Dubna, Russia

2. Automatic Equilibration and Movie Making, Workshop on High Throughput Molecular Dynamics, November 7-8th 2013, Barcelona, Spain

Posters

1. Investigating protein folding and binding. GRIB EXPO 2014, November 10th 2014, Barcelona, Spain

2. Kinetic, thermodynamic and structural characterization of the states of disordered proteins: a case study. CECAM Intrinsically Disordered Proteins: Connecting Computation, Physics and Biology, Sept. 1-5th 2013, Zurich, Switzerland. Designated as among top 5 posters of conference.

3. Network Component Characterization Using Molecular Dynamics simulations. Integrative Network Biology 2012: Network Medicine, May 11-13th 2012, Helsingor, Denmark

4. Visualizing rare biological events with Molecular Dynamics simulations. GRIB EXPO 2012, April 26th 2012, Barcelona, Spain

Chapter 7

APPENDIX: OTHER PUBLICATIONS

This section is to list work in which I contributed a minor part.

7.1 Dopamine transporter (DAT)

In collaboration with George Khelashivili & Harel Weinstein at Cornell University. Manuscript in preparation.

Summary

In this work we show that membrane lipids are directly involved in the function of the human dopamine transporter (hDAT).

ABSTRACT

Functional mechanisms of neurotransmitter:sodium symporters (NSS) proteins at a detailed molecular level have been primarily sought from studies of the cognate bacterial homolog, leucine transporter (LeuT), for which detailed structural knowledge is available. However, to what extent the inferences accumulated on LeuT can explain the corresponding functional mechanisms in the mammalian NSS transporters is not clear. Specifically, in contrast to LeuT, the NSS proteins contain functionally important long intracellular terminal domains whose mechanistic role in the function of these transporters has not been yet characterized on the molecular level. Here we provide, to our knowledge for the first time, investigation of the molecular mechanisms that involve the N-terminus of the human dopamine transporter (hDAT), in the functionally relevant state-to-state transitions in the transporter. The analysis of the extensive atomisic molecular dynamics (MD) simulations (totaling ~14μs of trajectories) of the full-length hDAT model in physiologically relevant lipid membranes, enriched in highly anionic phosphatidylinositol 4,5-biphosphate (PIP$_2$) lipids, revealed the outward-open to inward-open isomerization event in terms of rearrangements in specific structural motifs, previously identified in LeuT, that culminate in the destabilization and release of Na$^+$ ion from the functional Na2 site. But we also show that in hDAT this transition is related to PIP$_2$-mediated electrostatic association between the N-terminal domain and intracellular loop 4 (ICL4) segment of the transporter. We find that this association disrupts the intracellular gates that stabilize hDAT in the inward-closed state, and triggers, through allosteric coupling between the ICL4 and functionally relevant regions, including secondary substrate binding S2 site, large-scale concerted motions in the transporter that are similar to those underling outward-open to inward-open isomerization in LeuT. Presented data thus provide novel molecular-level mechanistic insights regarding the role of the N-terminal segment and PIP$_2$ lipids in functionally relevant dynamic transitions in the DAT and, with that, enhances our understanding of functional mechanisms of the NSS transporters.

Bibliography

[1] Nicholson RI, Gee JMW, Harper ME. EGFR and cancer prognosis. European Journal of Cancer. 2001 Sep;37, Supplement 4:9–15. Available from: `http://www.sciencedirect.com/science/article/pii/S0959804901002313`.

[2] Montagut C, Dalmases A, Bellosillo B, Crespo M, Pairet S, Iglesias M, et al. Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer. Nature Medicine. 2012 Feb;18(2):221–223. Available from: `http://www.nature.com/nm/journal/v18/n2/abs/nm.2609.html`.

[3] Wacker D, Wang C, Katritch V, Han GW, Huang XP, Vardy E, et al. Structural Features for Functional Selectivity at Serotonin Receptors. Science. 2013 May;340(6132):615–619. Available from: `http://www.sciencemag.org/content/340/6132/615`.

[4] Wang C, Jiang Y, Ma J, Wu H, Wacker D, Katritch V, et al. Structural Basis for Molecular Recognition at Serotonin Receptors. Science. 2013 May;340(6132):610–614. Available from: `http://www.sciencemag.org/content/340/6132/610`.

[5] Marti-Solano M, Iglesias A, Fabritiis Gd, Sanz F, Brea J, Loza MI, et al. Detection of New Biased Agonists for the Serotonin 5-HT2A Receptor: Modeling and Experimental Validation.

Molecular Pharmacology. 2015 Feb;p. mol.114.097022. Available from: `http://molpharm.aspetjournals.org/content/early/2015/02/06/mol.114.097022`.

[6] Kalyaanamoorthy S, Chen YPP. Modelling and enhanced molecular dynamics to steer structure-based drug discovery. Progress in Biophysics and Molecular Biology. 2014 May;114(3):123–136. Available from: `http://www.sciencedirect.com/science/article/pii/S0079610713000552`.

[7] Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Nature. 2007 Dec;450(7172):964–972. Available from: `http://www.nature.com/nature/journal/v450/n7172/full/nature06522.html`.

[8] Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature. 1958 Mar;181(4610):662–666.

[9] Editorial. Hard data. Nature. 2014 May;509(7500):260–260. Available from: `http://www.nature.com/news/hard-data-1.15216`.

[10] RCSB. PDB Current Holdings Breakdown; 2015. Available from: `http://www.rcsb.org/pdb/statistics/holdings.do`.

[11] Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. Biomolecular Simulation: A Computational Microscope for Molecular Biology. Annual Review of Biophysics. 2012;41(1):429–452. Available from: `http://www.annualreviews.org/doi/abs/10.1146/annurev-biophys-042910-155245`.

[12] Alber T, Gilbert WA, Ponzi DR, Petsko GA. The role of mobility in the substrate binding and catalytic machinery of enzymes. Ciba Foundation Symposium. 1983;93:4–24.

[13] Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, et al. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science (New York, NY). 1996 Mar;271(5253):1247–1254.

[14] Spolar RS, Record MT. Coupling of local folding to site-specific binding of proteins to DNA. Science. 1994 Feb;263(5148):777–784. Available from: http://www.sciencemag.org/content/263/5148/777.

[15] Aviles FJ, Chapman GE, Kneale GG, Crane-Robinson C, Bradbury EM. The Conformation of Histone H5. European Journal of Biochemistry. 1978 Aug;88(2):363–371. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1432-1033.1978.tb12457.x/abstract.

[16] Muchmore SW, Sattler M, Liang H, Meadows RP, Harlan JE, Yoon HS, et al. X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. Nature. 1996 May;381(6580):335–341.

[17] Hashem Y, Georges Ad, Fu J, Buss SN, Jossinet F, Jobe A, et al. High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome. Nature. 2013;Available from: http://www.nature.com/nature/journal/vaop/ncurrent/full/nature11872.html.

[18] Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature. 2013 May;497(7451):643–646. Available from: http://www.nature.com/nature/journal/v497/n7451/full/nature12162.html.

[19] Greber BJ, Boehringer D, Leibundgut M, Bieri P, Leitner A, Schmitz N, et al. The complete structure of the large subunit of the mammalian mitochondrial ribosome. Nature. 2014 Oct;advance online publication. Available from:

http://www.nature.com/nature/journal/vnfv/
ncurrent/full/nature13895.html.

[20] Plaschka C, Larivière L, Wenzeck L, Seizl M, Hemann M, Tegunov D, et al. Architecture of the RNA polymerase II-Mediator core initiation complex. Nature. 2015 Feb;advance online publication. Available from: http://www.nature.com/nature/journal/vaop/ncurrent/full/nature14229.html.

[21] Palmer III AG, Fairbrother WJ, Cavanagh J, Skelton NJ, Rance M. Protein NMR Spectroscopy, Second Edition: Principles and Practice. 2nd ed. Amsterdam ; Boston: Academic Press; 2005.

[22] Palmer III AG. Chemical exchange in biomacromolecules: Past, present, and future. Journal of Magnetic Resonance. 2014 Apr;241:3–17. Available from: http://www.sciencedirect.com/science/article/pii/S1090780714000159.

[23] Pellecchia M, Sem DS, Wüthrich K. Nmr in drug discovery. Nature Reviews Drug Discovery. 2002 Mar;1(3):211–219. Available from: http://www.nature.com/nrd/journal/v1/n3/full/nrd748.html.

[24] Kang C, Li Q. Solution NMR study of integral membrane proteins. Current Opinion in Chemical Biology. 2011 Aug;15(4):560–569. Available from: http://www.sciencedirect.com/science/article/pii/S1367593111000949.

[25] Ban D, Sabo TM, Griesinger C, Lee D. Measuring Dynamic and Kinetic Information in the Previously Inaccessible Supra-tc Window of Nanoseconds to Microseconds by Solution NMR Spectroscopy. Molecules. 2013 Sep;18(10):11904–11937. Available from: http://www.mdpi.com/1420-3049/18/10/11904.

[26] Opella SJ, Marassi FM. Structure Determination of Membrane Proteins by NMR Spectroscopy. Chemical Reviews. 2004;104(8):3587–3606. Available from: `http://dx.doi.org/10.1021/cr0304121`.

[27] PERMI P. Aspects of Coherence Transfer in High Molecular Weight Proteins. In: Webb GA, editor. Annual Reports on NMR Spectroscopy. vol. 53. Academic Press; 2004. p. 245–296. Available from: `http://www.sciencedirect.com/science/article/pii/S0066410304530053`.

[28] Tugarinov V, Hwang PM, Kay LE. Nuclear Magnetic Resonance Spectroscopy of High-Molecular-Weight Proteins. Annual Review of Biochemistry. 2004;73(1):107–146. Available from: `http://dx.doi.org/10.1146/annurev.biochem.73.011303.074004`.

[29] Pervushin K, Riek R, Wider G, Wüthrich K. Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proceedings of the National Academy of Sciences. 1997 Nov;94(23):12366–12371. Available from: `http://www.pnas.org/content/94/23/12366`.

[30] Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K. NMR analysis of a 900K GroEL–GroES complex. Nature. 2002 Jul;418(6894):207–211. Available from: `http://www.nature.com/nature/journal/v418/n6894/full/nature00860.html`.

[31] Roy R, Hohng S, Ha T. A practical guide to single-molecule FRET. Nature Methods. 2008 Jun;5(6):507–516. Available from: `http://www.nature.com/nmeth/journal/v5/n6/abs/nmeth.1208.html`.

[32] Haas E. The Study of Protein Folding and Dynamics by Determination of Intramolecular Distance Distributions and Their

Fluctuations Using Ensemble and Single-Molecule FRET Measurements. ChemPhysChem. 2005 May;6(5):858–870. Available from: `http://onlinelibrary.wiley.com/doi/10.1002/cphc.200400617/abstract`.

[33] Schuler B, Eaton WA. Protein folding studied by single-molecule FRET. Current Opinion in Structural Biology. 2008 Feb;18(1):16–26. Available from: `http://www.sciencedirect.com/science/article/pii/S0959440X07002035`.

[34] Plaxco KW, Millett IS, Segel DJ, Doniach S, Baker D. Chain collapse can occur concomitantly with the rate-limiting step in protein folding. Nature Structural & Molecular Biology. 1999 Jun;6(6):554–556. Available from: `http://www.nature.com/nsmb/journal/v6/n6/abs/nsb0699_554.html`.

[35] Hura GL, Menon AL, Hammel M, Rambo RP, Poole Ii FL, Tsutakawa SE, et al. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nature Methods. 2009 Aug;6(8):606–612. Available from: `http://www.nature.com/nmeth/journal/v6/n8/full/nmeth.1353.html`.

[36] Cavasotto CN, Phatak SS. Homology modeling in drug discovery: current trends and applications. Drug Discovery Today. 2009 Jul;14(13–14):676–683. Available from: `http://www.sciencedirect.com/science/article/pii/S1359644609001469`.

[37] Lexa KW, Carlson HA. Protein flexibility in docking and surface mapping. Quarterly Reviews of Biophysics. 2012 Aug;45(03):301–343. Available from: `http://journals.cambridge.org/article_S0033583512000066`.

[38] Fischer M, Coleman RG, Fraser JS, Shoichet BK. Incorporation of protein flexibility and conformational energy penal-

ties in docking screens to improve ligand discovery. Nature Chemistry. 2014 May;advance online publication. Available from: `http://www.nature.com/nchem/journal/vaop/ncurrent/abs/nchem.1954.html`.

[39] Binder K, Heermann D. Monte Carlo Simulation in Statistical Physics: An Introduction. 5th ed. Heidelberg ; New York: Springer; 2010.

[40] Rohl CA, Strauss CEM, Misura KMS, Baker D. Protein Structure Prediction Using Rosetta. In: Ludwig Brand and Michael L Johnson, editor. Methods in Enzymology. vol. Volume 383 of Numerical Computer Methods, Part D. Academic Press; 2004. p. 66–93. Available from: `http://www.sciencedirect.com/science/article/pii/S0076687904830040`.

[41] Das R, Baker D. Macromolecular Modeling with Rosetta. Annual Review of Biochemistry. 2008;77(1):363–382. Available from: `http://dx.doi.org/10.1146/annurev.biochem.77.062906.171838`.

[42] Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. Science. 2003 Nov;302(5649):1364–1368. Available from: `http://www.sciencemag.org/content/302/5649/1364`.

[43] Rajagopalan S, Wang C, Yu K, Kuzin AP, Richter F, Lew S, et al. Design of activated serine–containing catalytic triads with atomic-level accuracy. Nature Chemical Biology. 2014 Apr;advance online publication. Available from: `http://www.nature.com/nchembio/journal/vaop/ncurrent/full/nchembio.1498.html`.

[44] King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, André I, et al. Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. Science. 2012

Jun;336(6085):1171–1174. Available from: `http://www.sciencemag.org/content/336/6085/1171`.

[45] Kosloff R. Time-dependent quantum-mechanical methods for molecular dynamics. The Journal of Physical Chemistry. 1988;92(8):2087–2100. Available from: `http://dx.doi.org/10.1021/j100319a003`.

[46] Warshel A, Levitt M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. Journal of Molecular Biology. 1976 May;103(2):227–249. Available from: `http://www.sciencedirect.com/science/article/pii/0022283676903119`.

[47] Murphy RB, Philipp DM, Friesner RA. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. Journal of Computational Chemistry. 2000;21(16):1442–1457. Available from: `http://onlinelibrary.wiley.com/doi/10.1002/1096-987X(200012)21:16<1442::AID-JCC3>3.0.CO;2-O/abstract`.

[48] The Nobel Prize in Chemistry 2013 - Advanced Information; 2013. Available from: `http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/advanced.html`.

[49] Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nature Structural & Molecular Biology. 2002 Sep;9(9):646–652. Available from: `http://www.nature.com/nsmb/journal/v9/n9/full/nsb0902-646.html`.

[50] Rapaport DC. The Art of Molecular Dynamics Simulation. 2nd ed. Cambridge, UK ; New York, NY: Cambridge University Press; 2004.

[51] Mongan J, Case DA, McCammon JA. Constant pH molecular dynamics in generalized Born implicit solvent. Journal of Computational Chemistry. 2004;25(16):2038–2048. Available from: `http://onlinelibrary.wiley.com/doi/10.1002/jcc.20139/abstract`.

[52] Börjesson U, Hünenberger PH. Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines. The Journal of Chemical Physics. 2001 Jun;114(22):9706–9719. Available from: `http://scitation.aip.org/content/aip/journal/jcp/114/22/10.1063/1.1370959`.

[53] Donnini S, Tegeler F, Groenhof G, Grubmüller H. Constant pH Molecular Dynamics in Explicit Solvent with λ-Dynamics. Journal of Chemical Theory and Computation. 2011;7(6):1962–1978. Available from: `http://dx.doi.org/10.1021/ct200061r`.

[54] Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. BMC Biology. 2011 Oct;9(1):71. Available from: `http://www.biomedcentral.com/1741-7007/9/71/abstract`.

[55] Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. Journal of the American Chemical Society. 1995;117(19):5179–5197. Available from: `http://dx.doi.org/10.1021/ja00124a002`.

[56] Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. Journal of Computational Chemistry. 2003;24(16):1999–2012. Available from: `http://onlinelibrary.wiley.com/doi/10.1002/jcc.10349/abstract`.

[57] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. The Journal of Physical Chemistry B. 1998;102(18):3586–3616. Available from: http://dx.doi.org/10.1021/jp973084f.

[58] MacKerell AD, Banavali N, Foloppe N. Development and current status of the CHARMM force field for nucleic acids. Biopolymers. 2000;56(4):257–265. Available from: http://onlinelibrary.wiley.com/doi/10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W/abstract.

[59] Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. Journal of Computational Chemistry. 2004;25(9):1157–1174. Available from: http://onlinelibrary.wiley.com/doi/10.1002/jcc.20035/abstract.

[60] Mackerell AD, Feig M, Brooks CL. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. Journal of Computational Chemistry. 2004;25(11):1400–1415. Available from: http://onlinelibrary.wiley.com/doi/10.1002/jcc.20065/abstract.

[61] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of Computational Chemistry. 2010;31(4):671–690. Available from: http://onlinelibrary.wiley.com/doi/10.1002/jcc.21367/abstract.

[62] Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. Journal of Computational Chemistry. 2002 Dec;23(16):1623–1641.

[63] Best RB, Mittal J, Feig M, MacKerell Jr AD. Inclusion of Many-Body Effects in the Additive CHARMM Protein CMAP Potential Results in Enhanced Cooperativity of α-Helix and β-Hairpin Formation. Biophysical Journal. 2012 Sep;103(5):1045–1051. Available from: http://www.sciencedirect.com/science/article/pii/S0006349512008557.

[64] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Structure, Function, and Bioinformatics. 2006;65(3):712–725. Available from: http://onlinelibrary.wiley.com/doi/10.1002/prot.21123/abstract.

[65] Best RB, Buchete NV, Hummer G. Are Current Molecular Dynamics Force Fields too Helical? Biophysical Journal. 2008 Jan;95(1):L07–L09. Available from: http://www.cell.com/article/S0006349508702777/abstract.

[66] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins: Structure, Function, and Bioinformatics. 2010;78(8):1950–1958. Available from: http://onlinelibrary.wiley.com/doi/10.1002/prot.22711/abstract.

[67] Piana S, Lindorff-Larsen K, Shaw DE. How robust are protein folding simulations with respect to force field parameterization? Biophysical journal. 2011 May;100(9):L47–49.

[68] Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE. Systematic Validation of Protein Force Fields against

Experimental Data. PLoS ONE. 2012 Feb;7(2):e32131. Available from: `http://dx.doi.org/10.1371/journal.pone.0032131`.

[69] Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How Fast-Folding Proteins Fold. Science. 2011 Oct;334(6055):517–520. Available from: `http://www.sciencemag.org/content/334/6055/517`.

[70] Best RB, Zheng W, Mittal J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. Journal of Chemical Theory and Computation. 2014 Oct;Available from: `http://dx.doi.org/10.1021/ct500569b`.

[71] Skinner JJ, Yu W, Gichana EK, Baxa MC, Hinshaw JR, Freed KF, et al. Benchmarking all-atom simulations using hydrogen exchange. Proceedings of the National Academy of Sciences. 2014 Oct;p. 201404213. Available from: `http://www.pnas.org/content/early/2014/10/23/1404213111`.

[72] Mobley DL. Let's get honest about sampling. Journal of Computer-Aided Molecular Design. 2012 Jan;26(1):93–95. Available from: `http://link.springer.com/article/10.1007/s10822-011-9497-y`.

[73] Harvey MJ, De Fabritiis G. High-throughput molecular dynamics: the powerful new tool for drug discovery. Drug Discovery Today. 2012 Oct;17(19–20):1059–1062. Available from: `http://www.sciencedirect.com/science/article/pii/S1359644612001213`.

[74] Harvey MJ, De Fabritiis G. An Implementation of the Smooth Particle Mesh Ewald Method on GPU Hardware. Journal of Chemical Theory and Computation. 2009 Sep;5(9):2371–2377. Available from: `http://dx.doi.org/10.1021/ct900275y`.

[75] Harvey MJ, Giupponi G, Fabritiis GD. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. Journal of Chemical Theory and Computation. 2009 Jun;5(6):1632–1639. Available from: `http://dx.doi.org/10.1021/ct9000685`.

[76] Laio A, Parrinello M. Escaping free-energy minima. Proceedings of the National Academy of Sciences. 2002 Oct;99(20):12562–12566. Available from: `http://www.pnas.org/content/99/20/12562`.

[77] Dickson BM. Approaching a parameter-free metadynamics. Physical Review E. 2011 Sep;84(3):037701. Available from: `http://link.aps.org/doi/10.1103/PhysRevE.84.037701`.

[78] Limongelli V, Bonomi M, Parrinello M. Funnel metadynamics as accurate binding free-energy method. Proceedings of the National Academy of Sciences. 2013 Apr;Available from: `http://www.pnas.org/content/early/2013/04/03/1303186110`.

[79] Tiwary P, Limongelli V, Salvalaglio M, Parrinello M. Kinetics of protein–ligand unbinding: Predicting pathways, rates, and rate-limiting steps. Proceedings of the National Academy of Sciences. 2015 Feb;112(5):E386–E391. Available from: `http://www.pnas.org/content/112/5/E386`.

[80] Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. The Journal of Chemical Physics. 2004 Jun;120(24):11919–11929. Available from: `http://scitation.aip.org/content/aip/journal/jcp/120/24/10.1063/1.1755656`.

[81] Grant BJ, Gorfe AA, McCammon JA. Ras Conformational Switching: Simulating Nucleotide-Dependent Conformational Transitions with Accelerated Molecular Dynamics. PLoS Comput Biol.

2009 Mar;5(3):e1000325. Available from: `http://dx.doi.org/10.1371/journal.pcbi.1000325`.

[82] Hansen HS, Hünenberger PH. Using the local elevation method to construct optimized umbrella sampling potentials: Calculation of the relative free energies and interconversion barriers of glucopyranose ring conformers in water. Journal of Computational Chemistry. 2010;31(1):1–23. Available from: `http://onlinelibrary.wiley.com/doi/10.1002/jcc.21253/abstract`.

[83] Buch I, Sadiq SK, De Fabritiis G. Optimized Potential of Mean Force Calculations for Standard Binding Free Energies. Journal of Chemical Theory and Computation. 2011 Jun;7(6):1765–1772. Available from: `http://dx.doi.org/10.1021/ct2000638`.

[84] Zuckerman DM. Equilibrium Sampling in Biomolecular Simulations. Annual Review of Biophysics. 2011;40(1):41–62. Available from: `http://dx.doi.org/10.1146/annurev-biophys-042910-155255`.

[85] Chen J, Brooks CL, Khandogin J. Recent advances in implicit solvent-based methods for biomolecular simulations. Current Opinion in Structural Biology. 2008 Apr;18(2):140–148.

[86] Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI Coarse-Grained Force Field: Extension to Proteins. Journal of Chemical Theory and Computation. 2008;4(5):819–834. Available from: `http://dx.doi.org/10.1021/ct700324x`.

[87] de Jong DH, Singh G, Bennett WFD, Arnarez C, Wassenaar TA, Schäfer LV, et al. Improved Parameters for the Martini Coarse-Grained Protein Force Field. Journal of Chemical Theory and Computation. 2013 Jan;9(1):687–697. Available from: `http://dx.doi.org/10.1021/ct300646g`.

102

[88] Ingólfsson HI, Melo MN, van Eerden FJ, Arnarez C, Lopez CA, Wassenaar TA, et al. Lipid Organization of the Plasma Membrane. Journal of the American Chemical Society. 2014 Oct;136(41):14554–14559. Available from: `http://dx.doi.org/10.1021/ja507832e`.

[89] Moiset G, López CA, Bartelds R, Syga L, Rijpkema E, Cukkemane A, et al. Disaccharides Impact the Lateral Organization of Lipid Membranes. Journal of the American Chemical Society. 2014 Oct;Available from: `http://dx.doi.org/10.1021/ja505476c`.

[90] Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, et al. Anton, a special-purpose machine for molecular dynamics simulation. In: Proceedings of the 34th annual international symposium on Computer architecture. ISCA '07. New York, NY, USA: ACM; 2007. p. 1–12. Available from: `http://doi.acm.org/10.1145/1250662.1250664`.

[91] Shaw DE, Grossman JP, Bank JA, Batson B, Butts JA, Chao JC, et al. Anton 2: Raising the Bar for Performance and Programmability in a Special-purpose Molecular Dynamics Supercomputer. In: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. SC '14. Piscataway, NJ, USA: IEEE Press; 2014. p. 41–53. Available from: `http://dx.doi.org/10.1109/SC.2014.9`.

[92] Ohmura I, Morimoto G, Ohno Y, Hasegawa A, Taiji M. MDGRAPE-4: a special-purpose computer system for molecular dynamics simulations. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2014 Aug;372(2021):20130387. Available from: `http://rsta.royalsocietypublishing.org/content/372/2021/20130387`.

[93] Buch I, Harvey MJ, Giorgino T, Anderson DP, De Fabritiis G. High-Throughput All-Atom Molecular Dynamics Simulations Using Distributed Computing. Journal of Chemical Information and Modeling. 2010 Mar;50(3):397–403. Available from: `http://dx.doi.org/10.1021/ci900455r`.

[94] Singhal N, Snow CD, Pande VS. Using path sampling to build better Markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. The Journal of Chemical Physics. 2004 Jul;121(1):415–425. Available from: `http://scitation.aip.org/content/aip/journal/jcp/121/1/10.1063/1.1738647`.

[95] Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. The Journal of Chemical Physics. 2007 Apr;126(15):155101. Available from: `http://scitation.aip.org/content/aip/journal/jcp/126/15/10.1063/1.2714538`.

[96] Noé F, Fischer S. Transition networks for modeling the kinetics of conformational change in macromolecules. Current Opinion in Structural Biology. 2008 Apr;18(2):154–162. Available from: `http://www.sciencedirect.com/science/article/pii/S0959440X08000249`.

[97] Pande VS, Beauchamp K, Bowman GR. Everything you wanted to know about Markov State Models but were afraid to ask. Methods (San Diego, Calif). 2010 Jun;Available from: `http://www.ncbi.nlm.nih.gov/pubmed/20570730`.

[98] Prinz JH, Wu H, Sarich M, Keller B, Senne M, Held M, et al. Markov models of molecular kinetics: Generation and validation. The Journal of Chemical Physics. 2011 May;134(17):174105–174105–23. Available from:

`http://jcp.aip.org/resource/1/jcpsa6/v134/`
`i17/p174105_s1?isAuthorized=no.`

[99] Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. Proceedings of the National Academy of Sciences. 2009 Nov;106(45):19011–19016. Available from: `http://www.pnas.org/content/106/45/19011`.

[100] Voelz VA, Bowman GR, Beauchamp K, Pande VS. Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1-39). Journal of the American Chemical Society. 2010;132(5):1526–1528. Available from: `http://dx.doi.org/10.1021/ja9090353`.

[101] Radford IH, Fersht AR, Settanni G. Combination of Markov State Models and Kinetic Networks for the Analysis of Molecular Dynamics Simulations of Peptide Folding. The Journal of Physical Chemistry B. 2011;115(22):7459–7471. Available from: `http://dx.doi.org/10.1021/jp112158w`.

[102] Lane TJ, Bowman GR, Beauchamp K, Voelz VA, Pande VS. Markov State Model Reveals Folding and Functional Dynamics in Ultra-Long MD Trajectories. Journal of the American Chemical Society. 2011 Nov;133(45):18413–18419. Available from: `http://dx.doi.org/10.1021/ja207470h`.

[103] Buch I, Giorgino T, Fabritiis GD. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. Proceedings of the National Academy of Sciences. 2011 Jun;Available from: `http://www.pnas.org/content/early/2011/05/31/1103547108`.

[104] Held M, Metzner P, Prinz JH, Noé F. Mechanisms of Protein-Ligand Association and Its Modulation by Protein Mutations. Biophysical Journal. 2011 Feb;100(3):701–

710. Available from: `http://www.sciencedirect.com/science/article/pii/S000634951005215X`.

[105] Silva DA, Bowman GR, Sosa-Peinado A, Huang X. A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the LAO Protein. PLoS Comput Biol. 2011 May;7(5):e1002054. Available from: `http://dx.doi.org/10.1371/journal.pcbi.1002054`.

[106] Bowman GR, Beauchamp KA, Boxer G, Pande VS. Progress and challenges in the automated construction of Markov state models for full protein systems. The Journal of Chemical Physics. 2009 Sep;131(12):124101. Available from: `http://scitation.aip.org/content/aip/journal/jcp/131/12/10.1063/1.3216567`.

[107] Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F. Identification of slow molecular order parameters for Markov model construction. The Journal of Chemical Physics. 2013 Jul;139(1):015102–015102–13. Available from: `http://jcp.aip.org/resource/1/jcpsa6/v139/i1/p015102_s1`.

[108] Schwantes CR, Pande VS. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. Journal of Chemical Theory and Computation. 2013 Mar;Available from: `http://dx.doi.org/10.1021/ct300878a`.

[109] Doerr S, De Fabritiis G. On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations. Journal of Chemical Theory and Computation. 2014 Mar;Available from: `http://dx.doi.org/10.1021/ct400919u`.

[110] Fischer E. Einfluss der Configuration auf die Wirkung der Enzyme. Berichte der deutschen chemischen

Gesellschaft. 1894 Oct;27(3):2985–2993. Available from: `http://onlinelibrary.wiley.com/doi/10.1002/cber.18940270364/abstract`.

[111] Koshland DE. Application of a Theory of Enzyme Specificity to Protein Synthesis. Proceedings of the National Academy of Sciences of the United States of America. 1958 Feb;44(2):98–104.

[112] Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. Nature Chemical Biology. 2009 Nov;5(11):789–796. Available from: `http://www.nature.com/nchembio/journal/v5/n11/full/nchembio.232.html`.

[113] Vogt AD, Di Cera E. Conformational Selection or Induced Fit? A Critical Appraisal of the Kinetic Mechanism. Biochemistry. 2012 Jul;51(30):5894–5902. Available from: `http://dx.doi.org/10.1021/bi3006913`.

[114] Changeux JP, Edelstein S. Conformational selection or induced fit? 50 years of debate resolved. F1000 Biology Reports. 2011 Sep;3. Available from: `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3169905/`.

[115] Stein A, Rueda M, Panjkovich A, Orozco M, Aloy P. A Systematic Study of the Energetics Involved in Structural Changes upon Association and Connectivity in Protein Interaction Networks. Structure. 2011 Jun;19(6):881–889. Available from: `http://www.sciencedirect.com/science/article/pii/S096921261100133X`.

[116] Engelman DM. Membranes are more mosaic than fluid. Nature. 2005 Dec;438(7068):578–580. Available from: `http://www.nature.com/nature/journal/v438/n7068/abs/nature04394.html`.

[117] Lingwood D, Simons K. Lipid Rafts As a Membrane-Organizing Principle. Science. 2010 Jan;327(5961):46–50. Available from: http://www.sciencemag.org/content/327/5961/46.

[118] Daleke DL. Regulation of transbilayer plasma membrane phospholipid asymmetry. Journal of Lipid Research. 2003 Feb;44(2):233–242.

[119] Hannun YA, Obeid LM. Principles of bioactive lipid signalling: lessons from sphingolipids. Nature Reviews Molecular Cell Biology. 2008 Feb;9(2):139–150. Available from: http://www.nature.com/nrm/journal/v9/n2/abs/nrm2329.html.

[120] Wymann MP, Schneiter R. Lipid signalling in disease. Nature Reviews Molecular Cell Biology. 2008 Feb;9(2):162–176. Available from: http://www.nature.com/nrm/journal/v9/n2/full/nrm2335.html.

[121] Laganowsky A, Reading E, Allison TM, Ulmschneider MB, Degiacomi MT, Baldwin AJ, et al. Membrane proteins bind lipids selectively to modulate their structure and function. Nature. 2014 Jun;510(7503):172–175. Available from: http://www.nature.com/nature/journal/v510/n7503/full/nature13419.html.

[122] Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Journal of Molecular Biology. 2001 Jan;305(3):567–580. Available from: http://www.sciencedirect.com/science/article/pii/S0022283600943158.

[123] Yildirim MA, Goh KI, Cusick ME, Barabási AL, Vidal M. Drug—target network. Nature Biotechnology. 2007 Oct;25(10):1119–1126. Available from:

108

`http://www.nature.com/nbt/journal/v25/n10/`
`abs/nbt1338.html`.

[124] Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. Current Opinion in Structural Biology. 2008 Oct;18(5):581–586. Available from: `http://www.sciencedirect.com/science/article/pii/S0959440X08000997`.

[125] Hong M, Zhang Y, Hu F. Membrane Protein Structure and Dynamics from NMR Spectroscopy. Annual Review of Physical Chemistry. 2012;63(1):1–24. Available from: `http://dx.doi.org/10.1146/annurev-physchem-032511-143731`.

[126] Sanders CR, Sönnichsen F. Solution NMR of membrane proteins: practice and challenges. Magnetic Resonance in Chemistry. 2006;44(S1):S24–S40. Available from: `http://onlinelibrary.wiley.com/doi/10.1002/mrc.1816/abstract`.

[127] Bill RM, Henderson PJF, Iwata S, Kunji ERS, Michel H, Neutze R, et al. Overcoming barriers to membrane protein structure determination. Nature Biotechnology. 2011 Apr;29(4):335–340. Available from: `http://www.nature.com/nbt/journal/v29/n4/abs/nbt.1833.html`.

[128] Assadi-Porter FM, Tonelli M, Maillet E, Hallenga K, Benard O, Max M, et al. Direct NMR Detection of the Binding of Functional Ligands to the Human Sweet Receptor, a Heterodimeric Family 3 GPCR. Journal of the American Chemical Society. 2008;130(23):7212–7213. Available from: `http://dx.doi.org/10.1021/ja8016939`.

[129] Tapaneeyakorn S, Goddard AD, Oates J, Willis CL, Watts A. Solution- and solid-state NMR studies of GPCRs and their ligands. Biochimica et Biophysica Acta (BBA) -

Biomembranes. 2011 Jun;1808(6):1462–1475. Available from: `http://www.sciencedirect.com/science/article/pii/S0005273610003573`.

[130] Meyer B, Peters T. NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. Angewandte Chemie International Edition. 2003;42(8):864–890. Available from: `http://onlinelibrary.wiley.com/doi/10.1002/anie.200390233/abstract`.

[131] Liu JJ, Horst R, Katritch V, Stevens RC, Wüthrich K. Biased Signaling Pathways in β2-Adrenergic Receptor Characterized by 19F-NMR. Science. 2012 Mar;335(6072):1106–1110. Available from: `http://www.sciencemag.org/content/335/6072/1106`.

[132] Yakimchuk K. Protein Receptor-Ligand Interaction/Binding Assays. Materials and Methods. 2011 Jan;1. Available from: `http://www.labome.com/method/Protein-Receptor-Ligand-Interaction-Binding-Assays.html`.

[133] Dror RO, Pan AC, Arlow DH, Borhani DW, Maragakis P, Shan Y, et al. Pathway and mechanism of drug binding to G-protein-coupled receptors. Proceedings of the National Academy of Sciences. 2011 Aug;108(32):13118–13123. Available from: `http://www.pnas.org/content/108/32/13118`.

[134] Dror RO, Green HF, Valant C, Borhani DW, Valcourt JR, Pan AC, et al. Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. Nature. 2013 Oct;advance online publication. Available from: `http://www.nature.com/nature/journal/vaop/ncurrent/full/nature12595.html`.

[135] Nygaard R, Zou Y, Dror RO, Mildorf TJ, Arlow DH, Manglik A, et al. The Dynamic Process of β2-Adrenergic Receptor Activation. Cell. 2013 Jan;152(3):532–542. Available

from: http://www.sciencedirect.com/science/article/pii/S00928674413000111.

[136] Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, et al. Structure and dynamics of the M3 muscarinic acetylcholine receptor. Nature. 2012 Feb;482(7386):552–556. Available from: http://www.nature.com/nature/journal/v482/n7386/full/nature10867.html.

[137] Shan Y, Eastwood MP, Zhang X, Kim ET, Arkhipov A, Dror RO, et al. Oncogenic Mutations Counteract Intrinsic Disorder in the EGFR Kinase and Promote Receptor Dimerization. Cell. 2012 May;149(4):860–870. Available from: http://www.cell.com/abstract/S0092-8674(12)00416-3.

[138] Endres NF, Das R, Smith AW, Arkhipov A, Kovacs E, Huang Y, et al. Conformational Coupling across the Plasma Membrane in Activation of the EGF Receptor. Cell. 2013 Jan;152(3):543–556. Available from: http://www.sciencedirect.com/science/article/pii/S0092867412015541.

[139] Arkhipov A, Shan Y, Das R, Endres NF, Eastwood MP, Wemmer DE, et al. Architecture and Membrane Interactions of the EGF Receptor. Cell. 2013 Jan;152(3):557–569. Available from: http://www.cell.com/abstract/S0092-8674(12)01552-8.

[140] Jensen MØ, Borhani DW, Lindorff-Larsen K, Maragakis P, Jogini V, Eastwood MP, et al. Principles of conduction and hydrophobic gating in K+ channels. Proceedings of the National Academy of Sciences. 2010 Mar;107(13):5833–5838. Available from: http://www.pnas.org/content/107/13/5833.

[141] Jensen MØ, Jogini V, Borhani DW, Leffler AE, Dror RO, Shaw DE. Mechanism of Voltage Gating in Potassium Channels. Science. 2012 Apr;336(6078):229–233. Available from: http://www.sciencemag.org/content/336/6078/229.

[142] Ostmeyer J, Chakrapani S, Pan AC, Perozo E, Roux B. Recovery from slow inactivation in K+ channels is controlled by water molecules. Nature. 2013 Jul;advance online publication. Available from: `http://www.nature.com/nature/journal/vaop/ncurrent/full/nature12395.html?WT.ec_id=NATURE-20130801`.

[143] Mechoulam R, Parker LA. The Endocannabinoid System and the Brain. Annual Review of Psychology. 2013;64(1):21–47. Available from: `http://dx.doi.org/10.1146/annurev-psych-113011-143739`.

[144] Piomelli D. The molecular logic of endocannabinoid signalling. Nature Reviews Neuroscience. 2003 Nov;4(11):873–884. Available from: `http://www.nature.com/nrn/journal/v4/n11/abs/nrn1247.html`.

[145] Marzo VD, Bifulco M, Petrocellis LD. The endocannabinoid system and its therapeutic exploitation. Nature Reviews Drug Discovery. 2004 Sep;3(9):771–784. Available from: `http://www.nature.com/nrd/journal/v3/n9/full/nrd1495.html#top`.

[146] Mackie K. Cannabinoid Receptors as Therapeutic Targets. Annual Review of Pharmacology and Toxicology. 2006;46(1):101–122. Available from: `http://dx.doi.org/10.1146/annurev.pharmtox.46.120604.141254`.

[147] Pacher P, Bátkai S, Kunos G. The Endocannabinoid System as an Emerging Target of Pharmacotherapy. Pharmacological Reviews. 2006 Sep;58(3):389–462. Available from: `http://pharmrev.aspetjournals.org/content/58/3/389`.

[148] Christensen R, Kristensen PK, Bartels EM, Bliddal H, Astrup A. Efficacy and safety of the weight-loss drug rimonabant: a meta-analysis of randomised trials. The

Lancet. 2007 Nov;370(9600):1706–1713. Available from: http://www.sciencedirect.com/science/article/pii/S0140673607617218.

[149] McKinney MK, Cravatt BF. Structure and Function of Fatty Acid Amide Hydrolase. Annual Review of Biochemistry. 2005;74(1):411–432. Available from: http://dx.doi.org/10.1146/annurev.biochem.74.082803.133450.

[150] Dinh TP, Carpenter D, Leslie FM, Freund TF, Katona I, Sensi SL, et al. Brain monoglyceride lipase participating in endocannabinoid inactivation. Proceedings of the National Academy of Sciences. 2002 Aug;99(16):10819–10824. Available from: http://www.pnas.org/content/99/16/10819.

[151] Egertová M, Cravatt BF, Elphick MR. Comparative analysis of fatty acid amide hydrolase and cb(1) cannabinoid receptor expression in the mouse brain: evidence of a widespread role for fatty acid amide hydrolase in regulation of endocannabinoid signaling. Neuroscience. 2003;119(2):481–496.

[152] Venkatakrishnan AJ, Deupi X, Lebon G, Tate CG, Schertler GF, Babu MM. Molecular signatures of G-protein-coupled receptors. Nature. 2013 Feb;494(7436):185–194. Available from: http://www.nature.com/nature/journal/v494/n7436/full/nature11896.html.

[153] Hanson MA, Roth CB, Jo E, Griffith MT, Scott FL, Reinhart G, et al. Crystal Structure of a Lipid G Protein–Coupled Receptor. Science. 2012 Feb;335(6070):851–855. Available from: http://www.sciencemag.org/content/335/6070/851.

[154] Lee MJ, Thangada S, Claffey KP, Ancellin N, Liu CH, Kluk M, et al. Vascular endothelial cell adherens junction assembly and morphogenesis induced by sphingosine-1-phosphate. Cell. 1999 Oct;99(3):301–312.

[155] Garcia JG, Liu F, Verin AD, Birukova A, Dechert MA, Gerthoffer WT, et al. Sphingosine 1-phosphate promotes endothelial cell barrier integrity by Edg-dependent cytoskeletal rearrangement. The Journal of Clinical Investigation. 2001 Sep;108(5):689–701.

[156] Liu Y, Wada R, Yamashita T, Mi Y, Deng CX, Hobson JP, et al. Edg-1, the G protein-coupled receptor for sphingosine-1-phosphate, is essential for vascular maturation. The Journal of Clinical Investigation. 2000 Oct;106(8):951–961.

[157] Sharma N, Akhade AS, Qadri A. Sphingosine-1-phosphate suppresses TLR-induced CXCL8 secretion from human T cells. Journal of Leukocyte Biology. 2013 Apr;93(4):521–528.

[158] Sanchez T, Estrada-Hernandez T, Paik JH, Wu MT, Venkataraman K, Brinkmann V, et al. Phosphorylation and action of the immunomodulator FTY720 inhibits vascular endothelial cell growth factor-induced vascular permeability. The Journal of Biological Chemistry. 2003 Nov;278(47):47281–47290.

[159] Mullershausen F, Zecri F, Cetin C, Billich A, Guerini D, Seuwen K. Persistent signaling induced by FTY720-phosphate is mediated by internalized S1P1 receptors. Nature Chemical Biology. 2009 Jun;5(6):428–434. Available from: http://www.nature.com/nchembio/journal/v5/n6/abs/nchembio.173.html.

[160] Cannon RE, Peart JC, Hawkins BT, Campos CR, Miller DS. Targeting blood-brain barrier sphingolipid signaling reduces basal P-glycoprotein activity and improves drug delivery to the brain. Proceedings of the National Academy of Sciences. 2012 Sep;109(39):15930–15935. Available from: http://www.pnas.org/content/109/39/15930.

[161] Sadahira Y, Ruan F, Hakomori S, Igarashi Y. Sphingosine 1-phosphate, a specific endogenous signaling molecule controlling cell motility and tumor cell invasiveness. Proceedings

of the National Academy of Sciences. 1992 Oct;89(20):9686–9690. Available from: `http://www.pnas.org/content/89/20/9686`.

[162] Visentin B, Vekich JA, Sibbald BJ, Cavalli AL, Moreno KM, Matteo RG, et al. Validation of an anti-sphingosine-1-phosphate antibody as a potential therapeutic in reducing growth, invasion, and angiogenesis in multiple tumor lineages. Cancer Cell. 2006 Mar;9(3):225–238. Available from: `http://www.sciencedirect.com/science/article/pii/S1535610806000584`.

[163] Kim TH, Chung KY, Manglik A, Hansen AL, Dror RO, Mildorf TJ, et al. The Role of Ligands on the Equilibria Between Functional States of a G Protein-Coupled Receptor. Journal of the American Chemical Society. 2013 May;Available from: `http://dx.doi.org/10.1021/ja404305k`.

[164] Gautier A, Mott HR, Bostock MJ, Kirkpatrick JP, Nietlispach D. Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. Nature Structural & Molecular Biology. 2010 Jun;17(6):768–774. Available from: `http://www.nature.com/nsmb/journal/v17/n6/abs/nsmb.1807.html`.

[165] Park SH, Prytulla S, De Angelis AA, Brown JM, Kiefer H, Opella SJ. High-Resolution NMR Spectroscopy of a GPCR in Aligned Bicelles. Journal of the American Chemical Society. 2006;128(23):7402–7403. Available from: `http://dx.doi.org/10.1021/ja0606632`.

[166] Stanley N, Esteban-Martín S, De Fabritiis G. Kinetic modulation of a disordered protein domain by phosphorylation. Nature Communications. 2014 Oct;5(5272). Available from: `http://www.nature.com/ncomms/2014/141028/ncomms6272/full/ncomms6272.html`.

115

[167] Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. Journal of Molecular Graphics and Modelling. 2001 Feb;19(1):26–59. Available from: `http://www.sciencedirect.com/science/article/pii/S1093326300001388`.

[168] Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins: Structure, Function, and Bioinformatics. 2000 Nov;41(3):415–427. Available from: `http://onlinelibrary.wiley.com/doi/10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7/abstract`.

[169] Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. Journal of Molecular Biology. 1999 Oct;293(2):321–331. Available from: `http://www.sciencedirect.com/science/article/pii/S0022283699931108`.

[170] Tompa P. Structure and Function of Intrinsically Disordered Proteins. Chapman and Hall/CRC; 2009. Available from: `http://www.crcnetbase.com/isbn/9781420078930`.

[171] Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, et al. DisProt: the Database of Disordered Proteins. Nucleic Acids Research. 2007 Jan;35(suppl 1):D786–D793. Available from: `http://nar.oxfordjournals.org/content/35/suppl_1/D786`.

[172] Romero P, Obradovic Z, Kissinger C, Villafranca JE, Dunker AK. Identifying disordered regions in proteins from amino acid sequence. In: , International Conference on Neural Networks,1997. vol. 1; 1997. p. 90–95 vol.1.

[173] He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK. Predicting intrinsic disorder in proteins: an overview. Cell Research. 2009

Aug;19(8):929–949. Available from: `http://www.nature.com/cr/journal/v19/n8/full/cr200987a.html`.

[174] Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein Disorder Prediction: Implications for Structural Proteomics. Structure. 2003 Nov;11(11):1453–1459. Available from: `http://www.sciencedirect.com/science/article/pii/S0969212603002351`.

[175] Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 2010 Apr;1804(4):996–1010. Available from: `http://www.sciencedirect.com/science/article/pii/S1570963910000130`.

[176] Pancsa R, Tompa P. Structural Disorder in Eukaryotes. PLoS ONE. 2012 Apr;7(4):e34687. Available from: `http://dx.doi.org/10.1371/journal.pone.0034687`.

[177] Grotenbreg G, Ploegh H. Chemical biology: dressed-up proteins. Nature. 2007 Apr;446(7139):993–995.

[178] Geiss-Friedlander R, Melchior F. Concepts in sumoylation: a decade on. Nature reviews Molecular cell biology. 2007 Dec;8(12):947–956.

[179] Morrison RS, Kinoshita Y, Johnson MD, Uo T, Ho JT, McBee JK, et al. Proteomic analysis in the neurosciences. Molecular & cellular proteomics: MCP. 2002 Aug;1(8):553–560.

[180] Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, et al. Functional Anthology of Intrinsic Disorder. 3. Ligands, Post-Translational Modifications, and Diseases Associated with Intrinsically Disordered Proteins. Journal of Proteome Research. 2007 May;6(5):1917–1932. Available from: `http://dx.doi.org/10.1021/pr060394e`.

117

[181] Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, et al. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Research. 2004 Feb;32(3):1037–1049. Available from: `http://nar.oxfordjournals.org/content/32/3/1037`.

[182] Uversky VN, Oldfield CJ, Dunker AK. Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. Annual Review of Biophysics. 2008;37(1):215–246. Available from: `http://www.annualreviews.org/doi/abs/10.1146/annurev.biophys.37.032807.125924`.

[183] Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK. Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. Journal of Molecular Biology. 2002 Oct;323(3):573–584. Available from: `http://www.sciencedirect.com/science/article/pii/S0022283602009695`.

[184] Cheng Y, LeGall T, Oldfield CJ, Dunker AK, Uversky VN. Abundance of intrinsic disorder in protein associated with cardiovascular disease. Biochemistry. 2006 Sep;45(35):10448–10460.

[185] Lashuel HA, Hartley D, Petre BM, Walz T, Lansbury PT. Neurodegenerative disease: Amyloid pores from pathogenic mutations. Nature. 2002 Jul;418(6895):291–291. Available from: `http://www.nature.com/nature/journal/v418/n6895/abs/418291a.html`.

[186] Höppener JWM, Ahrén B, Lips CJM. Islet Amyloid and Type 2 Diabetes Mellitus. New England Journal of Medicine. 2000 Aug;343(6):411–419. Available from: `http://www.nejm.org/doi/full/10.1056/NEJM200008103430607`.

[187] Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z. Intrinsic disorder and protein function. Biochemistry. 2002 May;41(21):6573–6582.

118

[188] Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nature Reviews Molecular Cell Biology. 2005 Mar;6(3):197–208. Available from: `http://www.nature.com/nrm/journal/v6/n3/abs/nrm1589.html`.

[189] Dyson HJ. Expanding the proteome: disordered and alternatively folded proteins. Quarterly reviews of biophysics. 2011 Nov;44(4):467–518.

[190] Dogan J, Gianni S, Jemth P. The binding mechanisms of intrinsically disordered proteins. Physical Chemistry Chemical Physics. 2014 Mar;16(14):6323–6331. Available from: `http://pubs.rsc.org/en/content/articlelanding/2014/cp/c3cp54226b`.

[191] Huang Y, Liu Z. Do Intrinsically Disordered Proteins Possess High Specificity in Protein–Protein Interactions? Chemistry – A European Journal. 2013 Apr;19(14):4462–4467. Available from: `http://onlinelibrary.wiley.com/doi/10.1002/chem.201203100/abstract`.

[192] Espinoza-Fonseca LM. Reconciling binding mechanisms of intrinsically disordered proteins. Biochemical and Biophysical Research Communications. 2009 May;382(3):479–482. Available from: `http://www.sciencedirect.com/science/article/pii/S0006291X09004367`.

[193] Drobnak I, De Jonge N, Haesaerts S, Vesnaver G, Loris R, Lah J. Energetic Basis of Uncoupling Folding from Binding for an Intrinsically Disordered Protein. Journal of the American Chemical Society. 2013 Jan;135(4):1288–1294. Available from: `http://dx.doi.org/10.1021/ja305081b`.

[194] Hilser VJ, Thompson EB. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. Proceedings of the National Academy of Sciences. 2007 May;104(20):8311–8315. Available from: `http://www.pnas.org/content/104/20/8311`.

119

[195] Eliezer D. Biophysical characterization of intrinsically disordered proteins. Current Opinion in Structural Biology. 2009 Feb;19(1):23–30. Available from: `http://www.sciencedirect.com/science/article/pii/S0959440X08001784`.

[196] Snow CD, Nguyen H, Pande VS, Gruebele M. Absolute comparison of simulated and experimental protein-folding dynamics. Nature. 2002 Nov;420(6911):102–106. Available from: `http://www.nature.com/nature/journal/v420/n6911/full/nature01160.html`.

[197] Piana S, Lindorff-Larsen K, Shaw DE. Atomic-level description of ubiquitin folding. Proceedings of the National Academy of Sciences of the United States of America. 2013 Apr;110(15):5915–5920.

[198] Best RB, Hummer G, Eaton WA. Native contacts determine protein folding mechanisms in atomistic simulations. Proceedings of the National Academy of Sciences of the United States of America. 2013 Oct;110(44):17874–17879.

[199] Henry ER, Best RB, Eaton WA. Comparing a simple theoretical model for protein folding with all-atom molecular dynamics simulations. Proceedings of the National Academy of Sciences of the United States of America. 2013 Oct;110(44):17880–17885.

[200] Lin YS, Bowman GR, Beauchamp KA, Pande VS. Investigating How Peptide Length and a Pathogenic Mutation Modify the Structural Ensemble of Amyloid Beta Monomer. Biophysical Journal. 2012 Jan;102(2):315–324. Available from: `http://www.sciencedirect.com/science/article/pii/S0006349511053653`.

[201] Qiao Q, Bowman GR, Huang X. Dynamics of an intrinsically disordered protein reveal metastable conformations that potentially

seed aggregation. Journal of the American Chemical Society. 2013 Sep;Available from: `http://dx.doi.org/10.1021/ja403147m`.

[202] Chrivia JC, Kwok RPS, Lamb N, Hagiwara M, Montminy MR, Goodman RH. Phosphorylated CREB binds specifically to the nuclear protein CBP. Nature. 1993 Oct;365(6449):855–859. Available from: `http://www.nature.com/nature/journal/v365/n6449/abs/365855a0.html`.

[203] Radhakrishnan I, Pérez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. Cell. 1997 Dec;91(6):741–752.

[204] Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature. 2007 Jul;447(7147):1021–1025. Available from: `http://www.nature.com/nature/journal/v447/n7147/abs/nature05858.html`.

[205] Radhakrishnan I, Pérez-Alvarado GC, Dyson HJ, Wright PE. Conformational preferences in the Ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. FEBS letters. 1998 Jul;430(3):317–322.

[206] Ganguly D, Chen J. Atomistic details of the disordered states of KID and pKID. Implications in coupled binding and folding. Journal of the American Chemical Society. 2009 Apr;131(14):5214–5223.

[207] Chen HF. Molecular dynamics simulation of phosphorylated KID post-translational modification. PloS one. 2009;4(8):e6516.

[208] Turjanski AG, Gutkind JS, Best RB, Hummer G. Binding-induced folding of a natively unstructured transcription factor. PLoS computational biology. 2008 Apr;4(4):e1000060.

[209] Ganguly D, Chen J. Topology-based modeling of intrinsically disordered proteins: balancing intrinsic folding and intermolecular interactions. Proteins. 2011 Apr;79(4):1251–1266.

[210] Umezawa K, Ikebe J, Takano M, Nakamura H, Higo J. Conformational Ensembles of an Intrinsically Disordered Protein pKID with and without a KIX Domain in Explicit Solvent Investigated by All-Atom Multicanonical Molecular Dynamics. Biomolecules. 2012 Feb;2(1):104–121. Available from: `http://www.mdpi.com/2218-273X/2/1/104`.

[211] Mandell DJ, Chorny I, Groban ES, Wong SE, Levine E, Rapp CS, et al. Strengths of hydrogen bonds involving phosphorylated amino acid side chains. Journal of the American Chemical Society. 2007 Jan;129(4):820–827.

[212] Parker D, Jhala US, Radhakrishnan I, Yaffe MB, Reyes C, Shulman AI, et al. Analysis of an Activator:Coactivator Complex Reveals an Essential Role for Secondary Structure in Transcriptional Activation. Molecular Cell. 1998 Sep;2(3):353–359. Available from: `http://www.sciencedirect.com/science/article/pii/S1097276500802798`.

[213] Peterson CL, Laniel MA. Histones and histone modifications. Current Biology. 2004 Jul;14(14):R546–R551. Available from: `http://www.sciencedirect.com/science/article/pii/S0960982204004853`.

[214] Jiang ZG, McKnight CJ. A phosphorylation-induced conformation change in dematin headpiece. Structure (London, England: 1993). 2006 Feb;14(2):379–387.

[215] Volkman BF, Lipson D, Wemmer DE, Kern D. Two-state allosteric behavior in a single-domain signaling protein. Science (New York, NY). 2001 Mar;291(5512):2429–2433.

122

[216] Lubman OY, Waksman G. Dissection of the energetic coupling across the Src SH2 domain-tyrosyl phosphopeptide interface. Journal of molecular biology. 2002 Feb;316(2):291–304.

[217] Parrill AL, Wang Da, Bautista DL, Brocklyn JRV, Lorincz Z, Fischer DJ, et al. Identification of Edg1 Receptor Residues That Recognize Sphingosine 1-Phosphate. Journal of Biological Chemistry. 2000 Dec;275(50):39379–39384. Available from: `http://www.jbc.org/content/275/50/39379`.

[218] Borcherds W, Theillet FX, Katzer A, Finzel A, Mishall KM, Powell AT, et al. Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. Nature Chemical Biology. 2014 Nov;advance online publication. Available from: `http://www.nature.com/nchembio/journal/vaop/ncurrent/full/nchembio.1668.html`.

[219] Shan Y, Arkhipov A, Kim ET, Pan AC, Shaw DE. Transitions to catalytically inactive conformations in EGFR kinase. Proceedings of the National Academy of Sciences. 2013 Apr;Available from: `http://www.pnas.org/content/early/2013/03/28/1220843110`.

[220] Shammas SL, Travis AJ, Clarke J. Remarkably Fast Coupled Folding and Binding of the Intrinsically Disordered Transactivation Domain of cMyb to CBP KIX. The Journal of Physical Chemistry B. 2013 Oct;117(42):13346–13356. Available from: `http://dx.doi.org/10.1021/jp404267e`.