UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI CHIMICA FISICA ED ELETTROCHIMICA
SCUOLA DI DOTTORATO IN SCIENZE CHIMICHE
XXIII CICLO
CHIM02

Free energies in Biomolecular Simulations: From Protein-Protein Interactions to Unfolding Inhibition

Supervisor: Prof. Maurizio SIRONI

Ph.D. Candidate:
Giorgio SALADINO
Mat. Nr. R07829

A.A. 2009-2010

ai miei genitori

Il mondo è giunto all'orlo di un confine: se lo supererà, tutto potrà trasformarsi in follia, la gente andrà per le strade con una violetta in mano oppure si ammazzerà al primo incontro. E basterà poco, una goccia d'acqua, per far traboccare il bicchiere: magari che sulla strada ci sia una sola macchina, un solo uomo, o un solo decibel in più. C'è una specie di confine quantitativo che è vietato superare ma nessuno lo sorveglia e probabilmente nessuno sa che esiste.

Milan Kundera, ***L'Immortalità*** *(1990)*

Abstract

Free energies in Biomolecular Simulations:
From Protein-Protein Interactions To Unfolding Inhibition

Giorgio Saladino

supervisor *Prof. M. Sironi*

Dipartimento di Chimica Fisica ed Elettrochimica
Universita' degli Studi di Milano
via Golgi 19, 20133 Milano (Italy)

Part I - Microtubules are polymeric structures formed by the self association of tubulin dimers. They are extremely dynamical structures, that can undergo phases of growing and shrinking, playing a key role during cells proliferation process. Due to its importance for mitosis, tubulin is the target of many anticancer drugs currently in use or under clinical trial. The success of these molecules, however, is limited by the onset of resistant tumor cells during the treatment, so new resistance-proof compounds need to be developed. We analyze the protein-protein interactions allowing microtubules formation using molecular dynamics and free energy calculations. We were able to identify the most important amino acids for tubulin-tubulin binding and thus to design peptides, corresponding to tubulin subsequences. These peptides, able to interfere with microtubules formations, were proved to exhibit antitumoral activity.

Part II - Understanding the molecular mechanisms that allow some organisms to survive in extremely harsh conditions is an important achievement that might disclose a wide range of applications and that is constantly drawing the attention of many research fields. The simple small organic molecules, called osmolytes, responsible for the high adaptability of these living creatures are well known and of common use; nevertheless a full disclosure of the machinery behind their activity is still to be obtained. We developed a computational approach that, taking advantage of advanced simulation techniques, allowed to fully describe the effects of osmo-protectants on a small β-hairpin peptide and on a full mini-protein. The computational study allowed to highlight interesting new features and to develop a theory on the "osmoprotection driving force".

Contents

GENERAL INTRODUCTION AND THESIS OUTLINE

Free energies in Biomolecular Simulations

The evaluation of free energies from molecular simulations of biomolecules is extremely important for the purpose of relating microscopic data obtained "in silico" to measurable macroscopic quantities and to understand the physical and structural basis of biological phenomena. Ligand binding to receptors, protein-protein association, protein folding equilibria, transition between different conformations of DNA and transport of small molecules through channels are all processes whose understanding revolves around accurate free energy evaluations. However, despite being one of the most central quantities for biological processes, free energy is also one of the most difficult to compute from computer simulations. Indeed, free energy calculations require a widespread exploration of the configuration space that is extremely expensive, from a computational point of view, resulting generally unfeasible for biomolecules, whose underlying energy landscapes are rugged and complex. Many transformations of biological interest involve conformational changes that span time and/or length scales currently beyond the reach of typical molecular dynamics simulations. To overcome this limitation, many different methods have been developed to calculate free energy changes, whose effectiveness is generally limited to a specific problem. During the Ph.D. research activity, the attention has been focused on two rather challenging topics: the study of protein-protein interactions for drug development and the study of osmolytes effects on protein stability and

folding. Despite the dissimilarity of the systems studied, of the techniques employed and of the final purposes, we have repeatedly come upon the need for an effective strategy for free energy evaluation dealing with both topics, eventually leading to different tailored solutions.

Part I

Protein-Protein Interactions

INTRODUCTION

Protein-Protein Interactions in Drug Design

The discovery of a new drug is an extremely complex and multidisciplinary work, representing a constant challenge. New tools to address this problem are developed and improved at a fast pace, as quickly as less efficient old approaches fall into disuse. In this ever-changing scenario, the "high-throughput' techniques, that dominated the drug discovery scene in the last two decades, are now being questioned. The *in-vitro* testing of large libraries of compounds, extremely convenient when a molecule with a good activity is found, can quickly turn into a large waste of time, and money, when a so-called "no-hit" is produced, i.e. when the approach fails in identifying a good drug candidate. Moreover, many molecules, selected by *in-vitro* screening, can fail later as they progress through all the development phases. The inclusion of ADME (absorption, distribution, metabolism, excretion) tests early in the discovery process, in order to select "drug-like" compounds and minimize the failing chances, is becoming mandatory. Naturally, high-throughput ADME tests bear upon the already narrow advantage margin for these massive screening approaches. For these reasons computational approaches are becoming more and more popular as they provide a cost- and time-efficient way to pre-select "drug-like" compounds.

The computational paradigm

Computational approaches are now a key component of most drug discovery strategies, from hit identification to lead optimization and beyond, and approaches such as ligand- or structure-based virtual screening techniques are widely used in drug discovery. The docking of small molecules to protein binding sites, among the most crucial and popular "in silico" technique for drug design, was pioneered during the early 1980s, and remains one of the most used and constantly improved approaches. High-throughput docking is primarily used for hit-identification when only the structure of the receptor and of its binding site are known. However, similar calculations are often also used later on during lead optimization, when possible modifications can be tested "virtually" in a time- and cost-efficient way, before their synthesis. The docking process involves the prediction of ligand conformation and orientation (or posing) within a targeted binding site. Generally, docking studies are aimed at both an accurate structural modeling and a correct prediction of activity. However, understanding which features of a particular molecule are responsible for its activity and which modification might lead to an improvement of its potency is a challenging issue, and even more difficult is an accurate prediction from computer simulations.

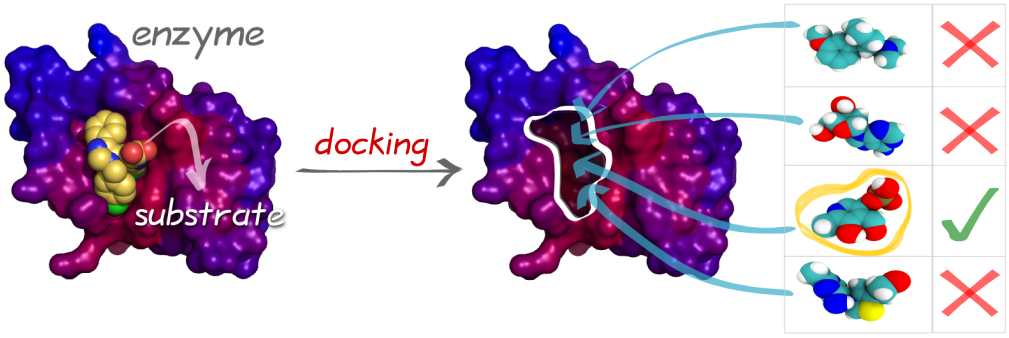

Figure I.1: The conventional strategy for computational drug design.

As a consequence, docking is generally applied iteratively in a multi-step process in which additional degrees of complexity are progressively introduced. The first step involves the use of docking algorithms in order to **pose** small molecules in the active site. Sampling must be performed with sufficient accuracy to identify the conformation that best matches the receptor structure, and must be fast enough to

permit the evaluation of thousands of compounds. Algorithms are complemented by **scoring functions** that are designed to predict the biological activity through the evaluation of interactions between compounds and potential targets. The most basic energy functions evaluate compounds affinity on the basis of shape and electrostatic complementarities. This relatively simple scoring functions continue to be heavily used, at least during the first stages of docking. Pre-selected conformers are often further tested using more complex scoring algorithms. The last step, consists in **ranking** the different tested compounds in order to selected the most suitable ones. "Posing, scoring and ranking" is the main principle, the paradigm, behind the so-called **Virtual Screening** (VS).

Innovative drug targets: protein-protein interactions

The presence of a well defined binding-site in the drug targets has always been desiderable, if not mandatory, for an efficient drug discovery project. It's no coincidence that, over the last few decades, most drugs that have been developed are enzyme inhibitors: enzymes bind naturally to small molecules, their substrate. This, indeed, highly simplifies the first phases of the development, offering the possibility of identifying, experimentally or computationally, small molecules that **mimic the substrate** and bind to these proteins, thus modulating their biological activity. The effectiveness of this strategy is also the main reason for the success of many computational approaches for rational drug design, based on docking. In fact, as the "posing, scoring and ranking" principle suggests, these methods heavily rely on the presence of a "cavity" in the protein surface for the pose of the different molecules. While many enzymes have still not been targeted, new opportunities are starting to be taken under consideration, among which inhibitors of protein-protein interactions (PPI) represent an attractive new class of molecules. Many proteins, including enzymes, exert part of their biological activity binding other proteins. Interfering with these interactions is a possible way of modulating the activity of such proteins. The structural diversity and large number of protein-protein interfaces offer an enormous amount of new targets for the pharmaceutical industry. However, protein interfaces have not evolved to bind to small molecules, as have enzymes and it may therefore be more difficult to identify inhibitors of protein-protein interactions. A second difficulty

comes from the diversity of protein interfaces. While large families of enzymes bind to the same substrate (for example, ATP for the kinases), and it is possible to use the knowledge gained for a particular member to more rapidly design compounds that target new members of the same family, this cannot be done for PPIs.

A strategy for the design of PPI-inhibiting drugs

Since protein interfaces involved in protein-protein association are not designed to bind small molecules, they lack binding sites, or cavities, for drug candidates. Considering the main idea behind conventional computational approaches, it is evident that the absence of a well-defined site in the target protein strongly affect these techniques, undermining their effectiveness. Indeed, the "posing, scoring and ranking" strategy is almost inapplicable when there is no site where to "pose" the molecules to be tested. New approaches are being developed to overcome this problem, some of which try to exploit features of PPIs themselves. It is well known, in fact, that while the contact surfaces involved in PPIs can be considerably wide, only some interfacial residues are actually important for the protein-protein association. These residues, called "hot-spots", are evenly distributed on the contact surfaces, strongly interacting with each other in order to form "bridges" between the two macromolecules.
A possible strategy is to replace the binding site used in conventional approaches with these hot-spots network, thus developing a drug able to bind to some of the network sites. While this idea is relatively straightforward from a theoretical point of view, its application might not be as easy as planned.

*In the following, a computational approach for an effective "in silico" design of PPIs inhibitors will be presented together with its first application to tubulin-tubulin interactions leading to microtubules formation. After a brief description of microtubules structure and biological function (**chapter 1**) and an overview of the computational strategy adopted (**chapter 2**), the results of our study will be presented, starting from the study of PPIs in a model system (**chapter 3**) and finishing with the analyses of the selected peptidic inhibitors (**chapter 4**) and experimental testing.*

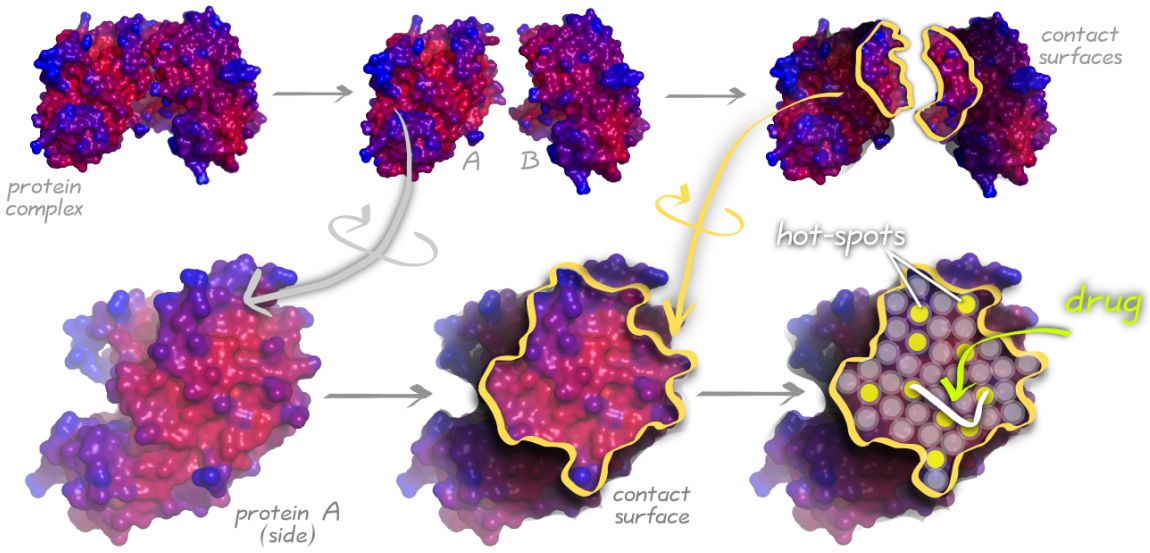

Figure I.2: The contact surfaces involved in PPI, while corrugated, lack binding sites for small molecules. Considering that only some residues are important for protein association (so-called "hot-spots") their interaction network can replace the conventional "cavity" as the drug design target, in order to obtain a inhibitor able to bind to some of the network sites.

CHAPTER 1

Microtubules

Microtubules (MTs) are important cytoskeletal structures that play a key role in several biological processes, from the formation of cilia and flagella to cell duplication. In particular, MTs form the so-called "mitotic spindle" that segregates chromosomes during mitosis, leading to the formation of two daughter cells from a mother one. If MTs are not able to subdivide the genetic material binding chromosomes, the cell is stuck in an intermediate stage of mitosis that, in most cases, degenerates into apoptosis (cells "programmed" death). Since MTs are crucial for mitosis, they are an ideal target for diseases characterized by an aberrant or unregulated cell duplication, as cancer. While different MT-targeting drugs exist, the problems connected to currently available molecules, including availability and the development of drug-resistance, fuel a constant search for new drugs. Since MTs formation and dynamical behavior is governed by tubulin self-association, the study of tubulin PPIs can a suitable way to develop a new drug.

1.1 Microtubules Structure and Dynamics

Microtubules are hollow cylindrical structures, usually with a 25 nm outer diameter, composed of a lattice of tubulin dimers. Tubulin is a hetero-dimeric protein, composed of an α and a β subunit, with a high degree of homology. Each subunit is a compact ellipsoid of approximate dimensions 46 x 40 x 65 Å with three domains: a N-terminal one hosting a GTP/GDP binding site, a central small one and a C-terminal prevalently helical domain. Tubulin α, β-dimers stack head-to-tail to form protofilaments that bind laterally to form sheets and are gradually rolled up into

a tube-like structure (Figure 1.1)[1]. MTs can be formed by a variable number of protofilaments from 9 to 16, but the majority of them is the results of the association of 13 protofilaments[2]. The final MT structure is organized in a polar manner such

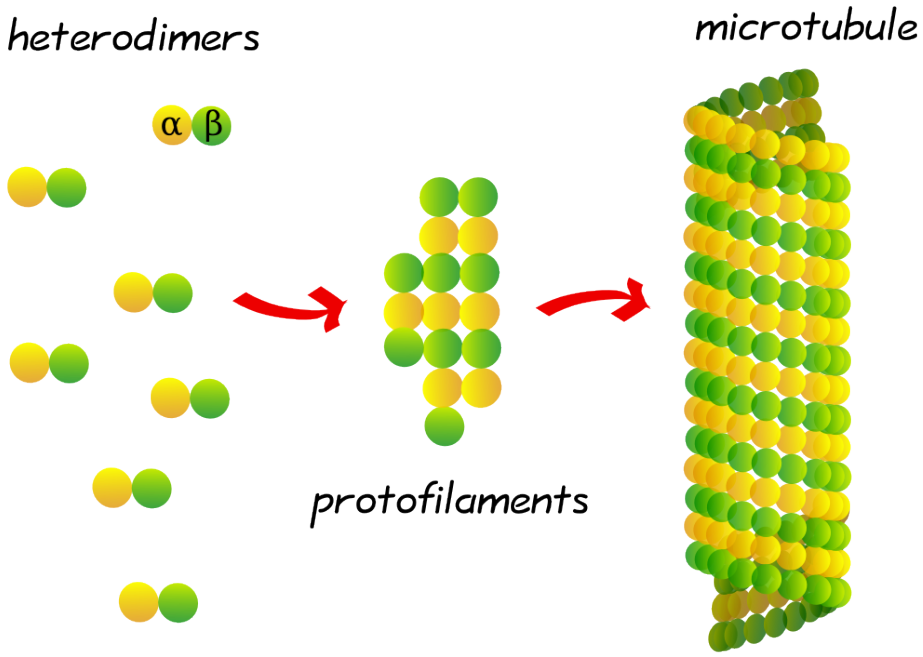

Figure 1.1: Tubulin α, β heterodimers polymerize to form protofilaments and microtubules.

that the α-tubulin subunit is exposed at one end, while the β-tubulin subunit is exposed at the other. The two ends of the MT are not equivalent: the β-capped one, called *plus-end*, is more dynamic, faster growing and faster shrinking than the other, called *minus-end*[3].

Microtubules polymerization is a complex mechanism, involving nucleation and subsequent elongation, fueled by the energy produced by the hydrolysis of GTP molecules bound to tubulin[4, 5]. One molecule of GTP must be associated with both the α and the β subunit for the association of heterodimers to occur. During the addition of a new subunit, the β-terminus of the MT interacts with the α subunit of the new dimer, promoting the hydrolysis of the GTP molecule bound to β-tubulin. The hydrolysis, however, takes place after a certain "lag" time from the addition of the dimer, in order to preserve a certain number of GTP-bound subunits at the terminus of the MT, the so-called *GTP-cap*[6]. After the lag period, the hydrolysis of GTP occurs: the β subunit becomes GDP-bound, while the α subunit continues to

bind a molecule of GTP (Figure 1.2).

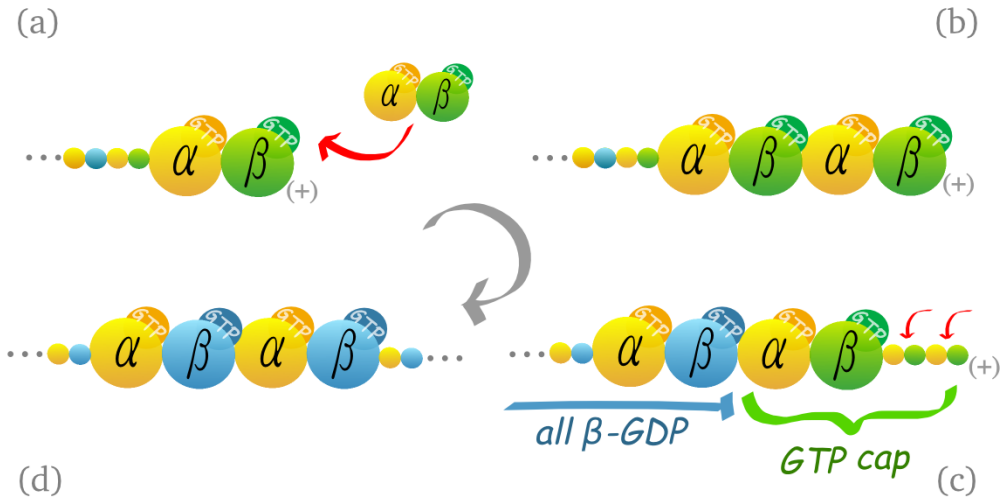

Figure 1.2: Subunits association to form the MT. (a) A new dimer, GTP-bound, can bind to the microtubule plus-end, i.e. the terminal β unit. (b) The "lag" between association and hydrolysis maintains a certain number of GTP-bound units at the plus-end. (c-d) the inner β units progressively hydrolyze the GTP molecule becoming GDP-bound, while α units remain GTP-bound.

1.1.1 Dynamic Instability

Even if MTs have a protective *GTP-cap* they are extremely dynamical structures that can undergo alternating phases of elongation and shortening, even at an overall steady state[7]. This behavior, termed "dynamic instability"[8], is unique among biopolymers. When the GDP-β-tubulin terminus is exposed, as a consequence of the stochastic loss of the *GTP-cap*, a conformational change is triggered, resulting in rapid depolymerization of the MT, an event called **catastrophe**[9]. On the other hand, the formation of a new layer of GTP-bound β-tubulins can restore the protective *GTP-cap* and switch off depolymerization, allowing the MT to grow again. This event is called **rescue**. Dynamic instability is characterized by long phases of slow progressive growth interrupted by "catastrophe-triggered" short phases of rapid shrinking. While *in vitro* the frequency of switching between elongation and depolymerization is regulated by tubulin concentration, *in vivo* it is influenced by the presence of different microtubules-binding proteins or molecules. Interestingly, even if the *minus-end* is less physiologically relevant, being generally bound to other proteins, it exhibit the same dynamic instability of the other end, despite lacking a "catastrophe-preventing"

structure as the *GTP-cap*[9].

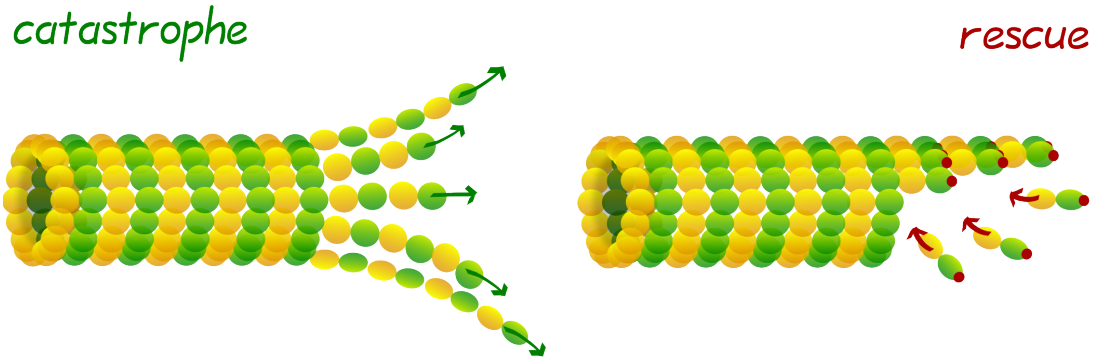

Figure 1.3: Microtubules can undergo alternating phases of elongation and shortening, a behavior termed "dynamic instability".

Molecular basis of Dynamic Instability

Dynamic instability is a non-equilibrium phenomenon and, consequently, it requires a source of energy to take place. The only possible source identified for tubulin is represented by the hydrolysis of the GTP molecules binding it. However, understanding how GTP affects the behavior of microtubules is still a matter of debate. One approach to understand the role of GTP is to investigate the polymerization/depolymerization dynamics in the presence of an a analogous compound, similar but non-hydrolysable, as *guanosine 5'-(β,γ - imidotriphosphate)*. The result of these studies[10, 11] revealed that the hydrolysis is not needed during the growth phase; tubulin polymerize effectively even with this GTP analogue, confirming that the energy of GTP hydrolysis is not necessary for this step. However, with the non-hydrolysable compound, the structure is more rigid and no sign of "dynamic instability" can be observed[4, 12]. Moreover, the MT lattice turned out to be substantially more stable in the presence of GTP-bound units only than in the presence of GDP[10]. This result suggests that the primary role of GTP hydrolysis is to destabilize the MT lattice, creating GDP-bound subunits that form weaker contacts. A proof of this hypothesis comes from studies done using a second analogue of GTP whose hydrolysis is possible under certain conditions. These studies have shown that, under normal conditions, disappearance of dynamic instability is observed, but the addition of hydrolysis-promoting agents (glycerol and Na^+ ions) triggers

a sudden depolymerization of the lattice[13]. The analyses also highlighted that GDP-

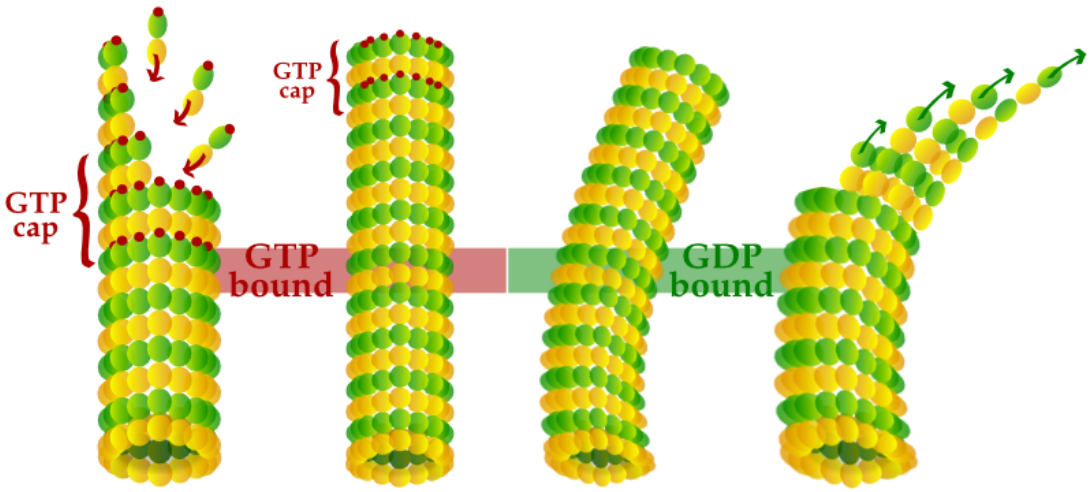

Figure 1.4: GTP-bound units force e "straight" conformation, while the loss of the GTP-cap allow GDP-bound units to adopt a more "bent" conformation that leads to depolymerization.

bound tubulins are more prone to adopt a curved conformation. Considering the results of these GTP studies, it had been postulated that the hydrolysis of GTP causes a structural transition leading to a "bent" conformation that destabilizes the microtubule[14, 15, 5]. The interaction between adjacent units (and the GTP-cap) prevents the adoption of a fully curved structure, storing the hydrolysis energy in the form of mechanical tension. When a GDP-bound tubulin unit remains exposed at the plus-end, as a consequence of the stochastic loss of the GTP-cap, this tension is released and the microtubule, force to assume a curved structure, depolymerizes rapidly.

1.2 Dynamic Instability as the key for Mitosis

Microtubules, as anticipated, play a crucial role in cell duplication, forming the so-called "mitotic spindle" that separates chromosomes[16].

The process of mitosis can be divided into phases according to the different organization and behavior of the chromosomes throughout the duplication.[17]. During *prophase* the chromosomes condense inside the cellular nucleus while microtubule nucleation is triggered at centrosomes and microtubules themselves become more

dynamic. The breakdown of the nuclear membrane marks the transition between prophase and *prometaphase*, during which the attachment of the microtubules to the chromosomes begins. During prometaphase, chromosomes exhibit a complex pattern of movements. Over time, these seemingly chaotic movements result in the *congression* of chromosomes to the spindle equator, such that more and more chromosomes become aligned between the separated spindle poles[18]. The less dynamic minus ends of MTs reside near the poles of the spindle, whereas the more dynamic plus ends extend towards the spindle equator. As a result, MTs between the spindle poles are organized into an antiparallel array, and microtubules outside of the spindle body form two radial asters that converge on the spindle poles. The MTs, due to their peculiar dynamics, can switch easily between phases of polymerization and depolymerization: while the minus-end remains anchored to the centrosome, the plus-end is able to explore the three-dimensional cellular space, until the microtubule is locked to the chromosome. This model, generally proposed for mitosis in vertebrates, is called "search- and-capture", and, in fact, is based on the search and subsequent capture of chromosomes by MTs within the cell space[19].

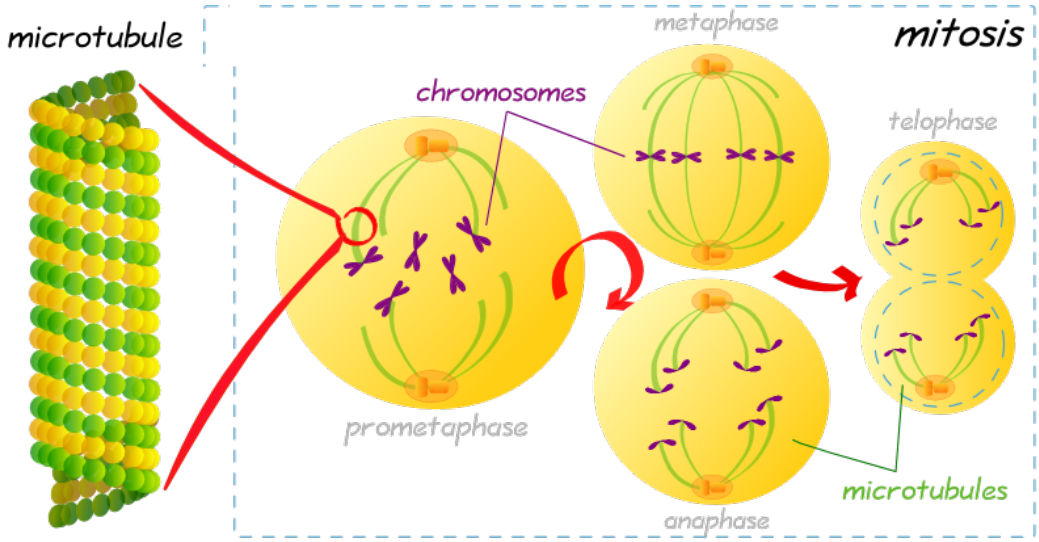

Figure 1.5: MTs play a crucial role during mitosis, forming the mitotic spindle and segregating chromosomes.

The key interaction between chromosomes and microtubules occurs at the centromere of the chromosome, which contains two identical macromolecular complexes termed *kinetochores*[20]. Progressively, spindle microtubules attach to each of the many chromosomes, positioning them at the spindle equator. The state of every chro-

mosome is monitored through a pathway termed the "spindle assembly checkpoint". Sporadic errors in chromosome attachment to the spindle microtubules are inevitable. Failure to correct these errors results in chromosome missegregation, a condition that underlies chromosomal instability, which is a hallmark of aggressive malignancies. How the spindle integrates all of its tasks remains unclear, even if significant progresses have been made recently in understanding the mechanisms that drive the complicated movements of chromosomes on the spindle. Congression of the last chromosome marks the transition to the next stage of mitosis, *metaphase*, the stage at which all chromosomes are aligned at the spindle equator. Shortly after metaphase alignment, the cohesion between sister chromatids is broken, and the cell enters *anaphase*. At this stage, the poleward movement of sister chromatids to the opposite poles of the spindle begins while poles themselves move further away from each other . During the next stage, *telophase*, the chromosomes decondense as the nuclear envelopes reform around the two daughter nuclei. The cell is finally divided in two by cytokinesis.

1.2.1 Microtubules-targeting Drugs

As already mentioned, tubulin and MTs crucial role in the cell duplication process is of particular interest for the development of anticancer drug; the study of MT-interacting molecules is therefore of primary importance for the development of new antitumoral strategies[21, 22]. MT-targeting compounds can be divided[23] in two main categories:

Microtubule destabilizing agents that inhibit the polymerization of MTs, including compounds as vinca alkaloids[24–26], colchicine[27] and combretastatin

Microtubule Stabilizing agents that enhance MTs polymerization; this class includes compounds as paclitaxel or docetaxel[28–30].

The main common feature that associate the first class with the second one is that, despite the mass of the microtubules is affected in a different way (increased or decreased), at concentrations 10-100 times lower, they are able to inhibit effectively the dynamics of MTs, stabilizing them kinetically[31, 32]. This ability is extremely interesting from a pharmacological point of view since a high concentrations of

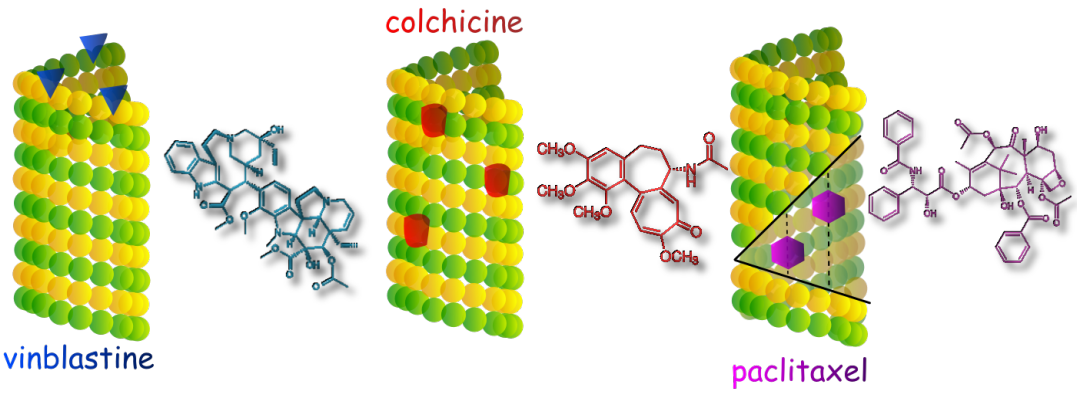

Figure 1.6: MT-targeting drugs as vinca alkaloids, colchicine or taxol

drug is necessary to influence the mass of microtubules, but acting on the dynamic aspects, it is possible to slow down, or even block, the metaphase-anaphase transition during mitosis, administering much smaller doses of drug. Whether the microtubule grows, whether it shrinks, it is the lack of a alternation between the two phases, resulting in the inhibition of "dynamic instability" which hinders mitosis, since the rapid disappearance of microtubule dynamics cannot ensure anymore the accurate and organized movement of chromosomes. All these compounds were mainly obtained by screening of natural compounds[33] and several problems arise from their limited availability. In addition, the complexity of some of these natural substances undermines the possibility of modifying these molecules to overcome drug resistance[34] when triggered.

1.2.2 The idea behind Rational Design

Despite the difficulties associated with the development of drugs targeting PPIs, due to the inherent features of protein-protein contact surfaces[35] and to their variability[36], a few encouraging results have already been obtained[37–39]. In particular PPIs inhibiting peptides were developed either by generating large libraries, for example exploiting the phage display technique[40, 41], or by selecting peptides corresponding to protein subsequences located at the protein-protein interface, both from structural[42, 43] and molecular modeling data[44, 45]. As already pointed out, only a small number of interfacial residues is fundamental for protein-protein association. As a consequence, if these "hot-spots" can be identified, the aforementioned approach, based on the development of peptides from protein subsequences, can be applied

effectively, selecting "hot-spots rich" subsequences. The idea behind our approach

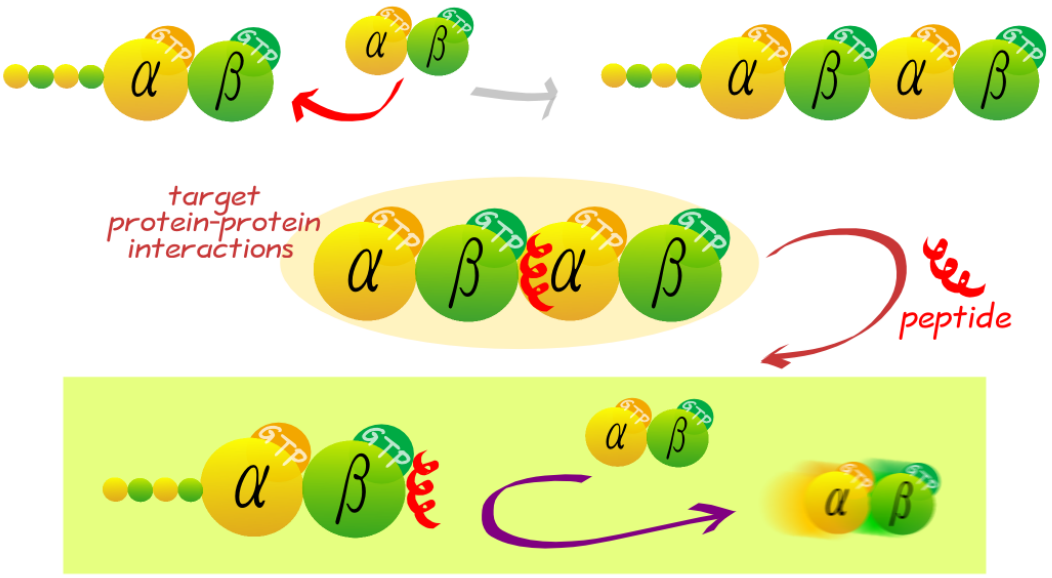

Figure 1.7: The main idea behind the project. If the highly-interacting "hot-spots" responsible for protein-protein binding are close enough, one may identify an "hot-stretch" (in red) that can be cut out to obtain a peptide. This one, binding to tubulin dimers, may block the MT, suppressing its dynamics.

is that, provided the highly-interacting "hot-spots" responsible for protein-protein binding are close enough, an "hot-stretch" (in red) might be identified, that can be cut out from the tubulin mother sequence to obtain a peptide. This one, binding to tubulin dimers, may interfere with MTs association, blocking them and suppressing their dynamics.

CHAPTER 2: COMPUTATIONAL METHODS I

Molecular Dynamics Simulations

In order to apply the approach exposed in the previous chapter (section 1.2.2) we need to perform a Molecular Dynamics (MD) simulation of a tubulin oligomer and identify the most important residues for protein-protein association. To do so, ***Computational Alanine Scanning*** (CAS) was selected as the most promising technique. Since CAS is based on binding free energy evaluations that, as anticipated, are a challenging task for computational chemistry, a theoretical framework for free energy calculations will be exposed in this chapter together with the description of the approaches used to calculate binding free energy and to perform CAS analyses. For completeness, a concise outline of the MD technique will be set above.

2.1 A Brief Overview of Molecular Dynamics

Computer simulations can provide important information on many phenomena thanks to their ability to describe the evolution of a system at an atomic level and on a time scale that is often inaccessible by other means. In particular, Molecular Dynamics (MD), despite the intrinsic approximations connected with a classical approach, is an effective and powerful tool to study large systems for which a full quantum level description would be unfeasible. The interest in the use of MD for the study of biomolecular systems, as proteins and macromolecules in general, is testified by the ever increasing number of published articles in this research field (Figure 2.1).
In a MD simulation, the system is described in the framework of Molecular Mechanics

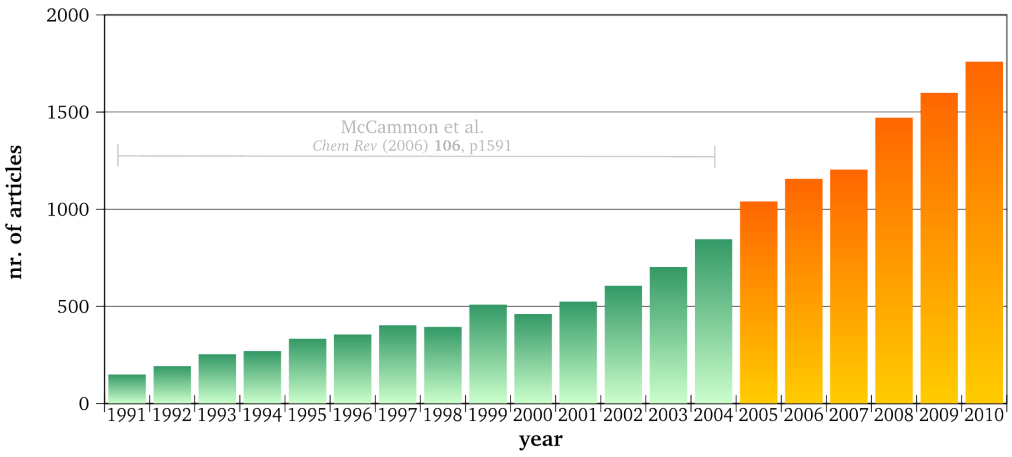

Figure 2.1: Published articles matching the query "protein AND (molecular dynamics)" in the ISI Web of Science database. The 1991-2004 data (green) have been reported in a popular review by McCammon et al.[46]: the outlined increasing trend has proved valid also in the last six years (orange).

(MM), where the atoms are represented as spherical masses connected by spring-like bonds[46]. All the forces are described using simple mathematical functions including several parameters, obtained by experiments or quantum mechanical studies and fit to reproduce experimental results. The set of functions together with the associated parameters is called a ***force-field***. The typical expression of a force-field includes contributions for bonds, angles, dihedrals and non-bonded interactions, as in the popular **AMBER** force-field

$$E_{\text{total}} = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{\text{non-bond}} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

In a MD simulation, the instantaneous forces of MM are calculated iteratively in order to evolve the system according to Newton's equations of motion. Several algorithms exist to integrate the equations of motion, such as ***Verlet algorithm***

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \frac{d^2x(t)}{dt^2}\Delta t^2$$

The integration step is governed by the fastest mode of the system, i.e. the vibration of bonds (especially those involving light H atoms), and is generally of the order of 1-2 fs, depending on the use of "bond-freezing" algorithms, as SHAKE or LINCS. In order to avoid "finite-size" effects, ***periodic boundary conditions*** are generally applied, while the non-bonded interactions are treated using a cut-off or specialized algorithms, as ***Particle Mesh Ewald*** (PME).(†)

2.2 Free Energy Calculations

As already mentioned, the calculation of free energy from a MD simulation is a challenging task. Indeed, free energy, entropy and related quantities (so-called "thermal quantities") are not simple averages of functions of the phase space coordinates. Rather, they are related to the volume in phase space that is accessible to the system. In other words, free energy directly depends on the system partition function and cannot be evaluated from a canonical average[47].
For a generic species i in solution the chemical potential is given by

$$\mu_{i,sol} = \mu^o_{i,sol} + RT \ln \frac{\gamma_i C_i}{C^o} \qquad [1]$$

where $\mu^o_{i,sol}$ is the standard chemical potential of i, C_i is the concentration, C^o the standard concentration (in the same units as C_i) and γ_i is the activity coefficient of i. According to statistical mechanics, the standard chemical potential can be written as

$$\mu^o_{i,sol} = -RT \ln \left(\frac{1}{V_{N,i}\, C^o} \frac{Q_{N,i}}{Q_{N,0}} \right) + P^o \bar{V}_i$$

where $Q_{N,i}$ is the canonical partition function for a system containing a large number N of solvent molecules and one solute molecule i and $Q_{N,0}$ is the canonical partition function of the N solvent molecules without the solute[49]. $\bar{V}_i$ is the partial molar volume of the solute at infinite dilution in the solvent (if $N \gg 1$, since $\bar{V}_i = V_{tot} - V_{N,0}$).

(†) This is intentionally a very concise summary of the basic idea behind MD. A more complete account can be found in various references[46–48].

Defining a proper laboratory frame, internal and external degrees of freedom can be separated for the solute i and the previous equation can be written as

$$\mu^o_{i,sol} = -RT \ln \left(Q^{rot}_{N,i} Q^{trasl}_{N,i} \frac{Z_{N,i}}{Z_{N,0}} \right) + P^o \bar{V}_i$$

and therefore

$$\mu^o_{i,sol} = -RT \ln \left(\frac{8\pi^2}{C^o} \prod_{i=1}^{M_i} (2\pi m_i RT)^{\frac{3}{2}} \frac{Z_{N,i}}{Z_{N,0}} \right) + P^o \bar{V}_i \qquad [2]$$

where

$$Z_{N,i} \equiv \int e^{-\beta U(r_i, r_s)} dr_i dr_s \qquad Z_{N,0} \equiv \int e^{-\beta U(r_s)} dr_s$$

$Z_{N,i}$ and $Z_{N,0}$ are called *configuration integrals* and explicitly contain the potential energy $U(r_i, r_s)$ that depends on the solute and solvent coordinates r_i and r_s. These integrals constitute the main difficulty in the calculation of free energies from a simulation, since their evaluation is connected to an efficient sampling of all the conformations involved in the process that, as anticipated, is beyond the reach of molecular dynamics[50]. Indeed, these terms, and free energy in general, depend not only on the lowest energy states, as other mechanical quantities, but on the volume of the conformational space itself.

2.2.1 Binding Free Energy

Considering the reaction

$$A + B \Leftrightarrow AB$$

where A and B can be a receptor and a ligand or two proteins involved in the formation of a protein complex AB. At equilibrium

$$\mu_{AB,sol} = \mu_{A,sol} + \mu_{B,sol}$$

and introducing the standard chemical potential in [1]

$$\mu^o_{A,sol} + RT \ln \frac{\gamma_A C_A}{C^o} + \mu^o_{B,sol} + RT \ln \frac{\gamma_B C_B}{C^o} = \mu^o_{AB,sol} + RT \ln \frac{\gamma_{AB} C_{AB}}{C^o}$$

The general expression for the standard free energy of binding is

$$\Delta G^o_{AB,sol} = \mu^o_{AB,sol} - \mu^o_{A,sol} - \mu^o_{B,sol} = -RT \ln \frac{\gamma_{AB} C_{AB} C^o}{\gamma_A C_A \gamma_B C_B} = -RT \ln K_{AB}$$

Since the kinetic contribution of each species have cancelled, using [2] for the standard chemical potential, this can be expressed as

$$\Delta G^o_{AB,sol} = -RT \ln \left(\frac{C^o}{8\pi^2} \right) \left(\frac{Z_{N,AB} Z_{N,0}}{Z_{N,A} Z_{N,B}} \right) + P^o \langle \Delta V_{AB} \rangle \qquad [3]$$

where the last term $P^o \langle \Delta V_{AB} \rangle$ is the pressure-volume work associated with changing the system size from the replacement of two free molecule (A, B) by one bound species (AB) and is generally considered to be negligibly small in water at 1 atm. This last term will be accordingly neglected in the following. Since the main term of ΔG^o_{AB} is the ratio of the configuration integrals, direct calculation of the binding energy from molecular dynamics simulations is unfeasible as well. A further simplification can be obtained using an implicit solvent. In this case we can write

$$\mu^o_{i,sol} = -RT \ln \left(\frac{8\pi^2}{C^o} \prod_{i=1}^{M_i} (2\pi m_i RT)^{\frac{3}{2}} \quad Z_i \right) + P^o \bar{V}_i$$

where

$$Z_i \equiv \frac{Z_{N,i}}{Z_{N,0}} = \int e^{-\beta[U(r_i) + W(r_i)]} dr_i$$

introducing the solvation energy as a function of the conformation

$$W(r_i) \equiv -RT \ln \left(\frac{\int e^{-\beta \Delta U(r_i, r_s)} e^{-\beta U(r_s)} dr_s}{\int e^{-\beta U(r_s)} dr_s} \right)$$

this term loosely corresponds to the work of transferring the solute in conformation r_i from gas phase to the solvent. Using implicit solvation, the final binding energy can be written as

$$\Delta G^o_{AB,sol} = -RT \ln \left(\frac{C^o}{8\pi^2} \right) \left(\frac{Z_{AB}}{Z_A Z_B} \right) \qquad [4]$$

Many computational techniques have been developed to accurately estimate the free energy and binding free energy in particular. The most rigorous of them are based on so-called "alchemical" or structural transformations, such as ***Free Energy Perturbation*** (FEP)[51] or ***Thermodynamic Integration*** (TI)[52]. As it will be extensively discussed in the **Computational Methods II** section, several simulation techniques based on MD, as ***Replica Exchange*** (RE) or ***metadynamics***, have been developed to sample the conformational space in a more effective way. Since exhaustive phase space sampling makes it possible to compute relative free energies directly, these techniques provide an alternative way to overcome the problem of free energy calculations.

End-points methods

So-called ***End-points*** methods evaluate only the initial and final state of the system. Since they are less computationally expensive, they provide an alternative to the more expensive techniques and are suitable for a greater variety of applications. These approaches are generally based on a decomposition of the total free energy into different contributions, both enthalpic and entropic. Typically, these approaches manipulate [4] into the form

$$\Delta G^o_{AB,sol} = \langle U_{AB} \rangle - \langle U_A \rangle - \langle U_B \rangle + \langle W_{AB} \rangle - \langle W_A \rangle - \langle W_B \rangle - T\Delta S^o$$

where U_X and W_X are the (Boltzmann-averaged) potential and solvation energy of species X. More concisely

$$\Delta G^{o}_{AB,sol} = \langle \Delta U + \Delta W \rangle - T\Delta S^{o} \quad {}^{(*)} \qquad [5]$$

Molecular Mechanics Generalized Born Surface Area (MM-GBSA), that takes its name from the different approaches used to calculate the different contributions, is among the most used methods of this kind.

2.2.2 The MM-GBSA Approach

The MM-GBSA approach[53] uses a thermodynamic cycle (Figure 2.2) to calculate the binding free energy on the basis of [5]. The cycle can be connected to [5] using the

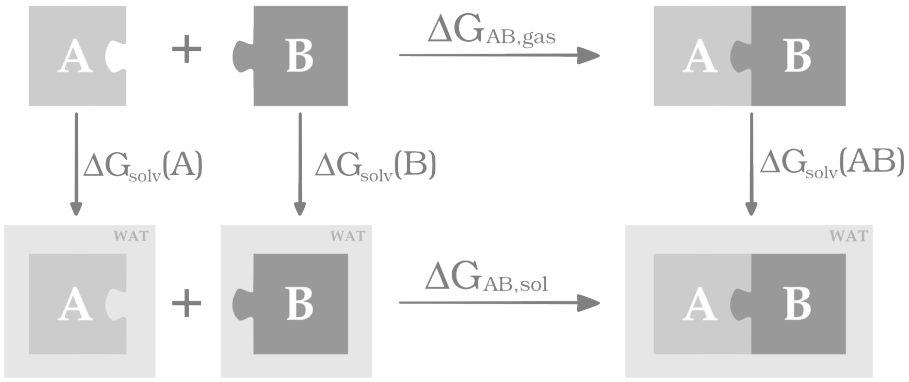

Figure 2.2: The thermodynamic cycle used to calculate the binding free energy in solution $\Delta G_{AB,sol}$. The binding energy is decomposed into a vacuum binding free energy $\Delta G_{AB,gas}$ and solvation free energies of all the species.

proposed working definition of $W(X)$, i.e. the work of transferring X from gas phase to the solvent. Indeed, the solvation contribution can be separated considering the reaction

$$X_{gas} \xrightarrow{\Delta G_{solv}(X)} X_{sol}$$

(*) It is to note that, since solvation energies are actually free energies, they include an entropic component and the entropic term ΔS^{o} reported in [5] is actually only the configuration entropy change ΔS^{o}_{config}, i.e. the change in entropy associated with the different species motions upon binding.

where $\Delta G_{solv}(X)$ is the free energy of solvation of X

$$\Delta G_{solv}(X) = G(X)_{sol} - G(X)_{gas}$$

the free energy of X in the solvent is then

$$G(X)_{sol} = G(X)_{gas} + \Delta G_{solv}(X)$$

therefore the binding energy [(†)]

$$\Delta G_{AB,sol} = G(AB)_{sol} - G(A)_{sol} - G(B)_{sol}$$

is written as

$$\Delta G_{AB,sol} = \Delta G_{AB,gas} + \underbrace{\Delta G_{solv}(AB) - \Delta G_{solv}(A) - \Delta G_{solv}(B)}_{\Delta\Delta G_{solv}}$$

decomposing $\Delta G_{AB,sol}$ into a vacuum term $\Delta G_{AB,gas}$ and a solvation term, according to the cycle (Figure 2.2). The separation of the two term is fundamental and allows to calculate the solvation term using implicit solvent models. Theoretically, $\Delta G_{AB,sol}$ can be calculated without the thermodynamic cycle and using an explicit solvent, however the most prominent contributions would result from interactions between the many solvent molecule, with consequent large fluctuations on the final ΔG value (likely to be orders of magnitude larger than ΔG itself). Considering that

$$G(X)_{sol} = U(X)_{gas} - TS(X)_{gas} + \Delta G_{solv}(X) \qquad [6]$$

it can be easily observed that, as compared with [5], $\Delta G_{AB,gas}$ loosely corresponds to $\langle \Delta U \rangle - T\Delta S$ while $\Delta\Delta G_{solv}$ corresponds to $\langle \Delta W \rangle$. To calculate the $G(X)_{sol}$ this equation is generally written as

[(†)] note that the standard sign in ΔG^o is dropped since standard concentration is 1 M or 1 molecule/1660 Å^3, corresponding to one molecule in a ∼12 Å-edge box, which is not generally ensured.

$$G(X)_{sol} = \langle E_{MM} \rangle_X + \Delta G_{solv}(X)$$

and therefore

$$\Delta G_{AB,sol} = \langle E_{MM} \rangle_{AB} - \langle E_{MM} \rangle_A - \langle E_{MM} \rangle_B + \Delta\Delta G_{solv}$$

$$= \langle \Delta E_{MM} \rangle + \Delta\Delta G_{solv}$$

where $\langle E_{MM} \rangle_X$ is the average Molecular Mechanics (MM) energy calculated during the MD simulation(‖)

$$\Delta E_{MM} = \underbrace{\Delta E_{bonds} + \Delta E_{angles} + \Delta E_{dihedrals}}_{\Delta E_{intra}} + \underbrace{\Delta E_{ele} + \Delta E_{vdw}}_{\Delta E_{inter}}$$

The Solvation Term

While ΔE_{MM} can be easily calculated from a MD simulation in explicit solvent without any further effort, the solvation term must be evaluated "a posteriori" using implicit solvent calculations(‡) The solvation term $\Delta\Delta G_{solv}$ can be decomposed into an electrostatic term and a non-electrostatic term

$$\Delta\Delta G_{solv} = \Delta\Delta G_{elec} + \Delta\Delta G_{non-elec} \qquad [7]$$

$$= \Delta G_{elec} + \Delta G_{vdw} + \Delta G_{cav}$$

(‖) Rigorously $\langle E_{MM} \rangle_X$ should contain also the configurational entropy according to [5] and [6], see section 2.2.2 on page 29 for details.

(‡) $\Delta\Delta G_{solv}$, as explained in the previous section, must be calculated using an implicit solvent model to avoid large fluctuations in ΔG. On the other hand, MD simulation can be carried out in explicit or implicit solvent. Explicit solvation will be assumed in the following.

where the non-electrostatic term has been further decomposed into a Van der Waals term and a "cavity" term, that corresponds to the energy needed to create the cavity occupied by the solute in the solvent.

The first term, the electrostatic one, can be easily computed using the Generalized Born (GB) model[54]. In 1920, Max Born derived the expression for the electrostatic component of the solvation energy associated with the creation of a charge q_i in a spherical cavity with radius a_i in a solvent with dielectric constant ϵ

$$\Delta G_{elec} = -\frac{1}{2}\frac{{q_i}^2}{a_i}\left(1 - \frac{1}{\epsilon}\right)$$

therefore, for N charges q_i the equation will be

$$\Delta G_{elec} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{i}^{N}\frac{{q_i}^2}{a_i} \qquad [8]$$

to which an additional term, accounting for the coulombian mutual interaction of the charges, when passing from gas phase to the solvent, must be added

$$\underbrace{\frac{1}{2}\sum_{i,j(i\neq j)}^{N}\frac{q_i q_j}{\epsilon r_{ij}}}_{solv}\underbrace{-\frac{1}{2}\sum_{i,j(i\neq j)}^{N}\frac{q_i q_j}{r_{ij}}}_{gas} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{i,j(i\neq j)}^{N}\frac{q_i q_j}{r_{ij}}$$

yielding the general equation

$$\Delta G_{elec} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{i,j(i\neq j)}^{N}\frac{q_i q_j}{r_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{i}^{N}\frac{{q_i}^2}{a_i} \qquad [9]$$

However, in a molecule, or in a chemical system in general, we don't have only isolated charges (as ions), but mostly atoms with atomic charges that are connected by bonds. Therefore [9] must be modified to take into account this feature of chemical systems

$$\Delta G_{GB} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right) \sum_{i,j(i \neq j)}^{N} \frac{q_i q_j}{f\left(r_{ij}, a_{ij}\right)}$$

yielding the Generalized-Born equation, where

$$f\left(r_{ij}, a_{ij}\right) = \sqrt{{r_{ij}}^2 + {a_{ij}}^2 e^{-D}} \qquad a_{ij} = \sqrt{a_i a_j} \qquad D = \frac{{r_{ij}}^2}{\left(2a_{ij}\right)^2}$$

This equation equals [9] if the charges are well separated since if $r_{ij} >> a_{ij}$

$$D = \frac{{r_{ij}}^2}{\left(2a_{ij}\right)^2} >> 1 \quad \rightarrow \quad e^{-D} \approx 0 \quad \rightarrow \quad f\left(r_{ij}, a_{ij}\right) = r_{ij}$$

alternatively, if $i = j(r_{ij} = 0)$

$$a_{ij} = a_{ii} \quad \rightarrow \quad r_{ij} = 0 \ , \ D = 0 \quad \rightarrow \quad f\left(0, a_{ii}\right) = \sqrt{0 + {a_i}^2 e^0} = a_i$$

and we obtain [8].

The second term, the non-electrostatic contribution

$$\Delta G_{nelec} = \Delta G_{cav} + \Delta G_{vdw}$$

includes a Van der Waals term and a cavity term, ΔG_{cav}, that is substantially the work to create a cavity for the solute in the solvent minus the entropy needed to reorganize the solvent molecules around the solute. Since both the reorganization of the solvent and the Van der Waals term are significant only for the first solvation shells, ΔG_{cav} and ΔG_{vdw} can be considered proportional to the surface of the solute exposed to the solvent. For this reason, $\Delta G_{non-elec}$ is generally calculated according to the ***(Solvent Accessible) Surface Area*** (SA) approach and, in particular, as

$$\Delta G_{SA} = \Delta G_{cav} + \Delta G_{vdw} = \gamma A + b \qquad [10]$$

where A is the total solvent accessible SA, while γ and B are two constants.

The Entropic Term

So far, we have omitted the entropic term in the description of the MM-GBSA approach. However, as explained in section 2.2.1 on page 23 the entropy of solvation is included in the calculated solvation term. Hence, of the total entropy

$$\Delta S = \Delta S_{config} + \Delta S_{solv}$$

only the first term is missing. While for some purposes (as Computational Alanine Scanning, see section 2.3 on page 30) this can be neglected, ΔS_{config} contribution is mandatory to achieve accurate binding energies. Unfortunately, even if the use of end-points approaches allows to separate the free energy in different contributions, avoiding the direct calculation of configurational integrals, ΔS_{config} remains deeply connected to them and difficult to calculate[47, 50]. Generally, this term can be written as

$$\Delta S_{config} = \Delta S_{ext} + \Delta S_{int}$$

considering external and internal degrees of freedom, corresponding

$$\Delta S_{config} = \Delta S_{rot+trans} + \Delta S_{vib}$$

where the two terms have been identified as the contribution due to the loss of translational and rotational freedom and to the internal vibration(‡)
While $\Delta S_{rot+trans}$ might prevail on ΔS_{vib} in the case of small molecules binding to large receptors, this term is generally neglected for protein-protein association, where the change in vibrational entropy is supposed to be more relevant. ΔS_{vib} can be calculated using normal mode analysis that, however, is too demanding to be applied to an entire protein-protein complex. For all these reason, the inclusion of entropic

(‡) The nature of these contributions and their relative importance is still a matter of debate.

contributions in free energy calculations is still heterogeneous and, mostly, limited to some approximated terms. In the present work, the inclusion of entropic terms will be limited, especially considering that these terms can be neglected in Computational Alanine Scanning calculations, that are our final purpose.

Application of MM-GBSA

Combining the different contributions, the $\Delta G_{AB,sol}$ is calculated as

$$\Delta G_{AB,sol} = \langle \Delta E_{MM} \rangle + \Delta G_{GB} + \Delta G_{SA}$$

the main equation behind the MM-GBSA approach. There is no univocal way to proceed in order to apply this equation. The MD simulations can be carried out in implicit or explicit solvent and, in addition, since we need the different contributions for each species (A, B and the complex AB), distinct simulations of all the species may be performed, or alternatively a simulation of the complex only. While the choice of an implicit solvent might speed up the simulation time, explicit solvation has proved to be a better approximation for biomolecular processes and in particular for protein association. Concerning the simulation instead, even if performing three different MDs for A, B and AB is, theoretically, a better solution, thanks to an advantageous compensation of different errors, the use of a single simulation of the complex yield better results. Calculations for A and B alone are performed "a posteriori" removing one of the two species from the trajectory of the AB complex.

2.3 CAS - Computational Alanine Scanning

Alanine Scanning (AS) mutagenesis is a valuable procedure to identify important residues ("hot-spots") at the protein-protein interface[55]. It involves the mutation of a generic residue X into an alanine, in order to evaluate the difference in binding energy upon mutation

$$\Delta\Delta G_X = \Delta G_{mut,X} - \Delta G_{wild}$$

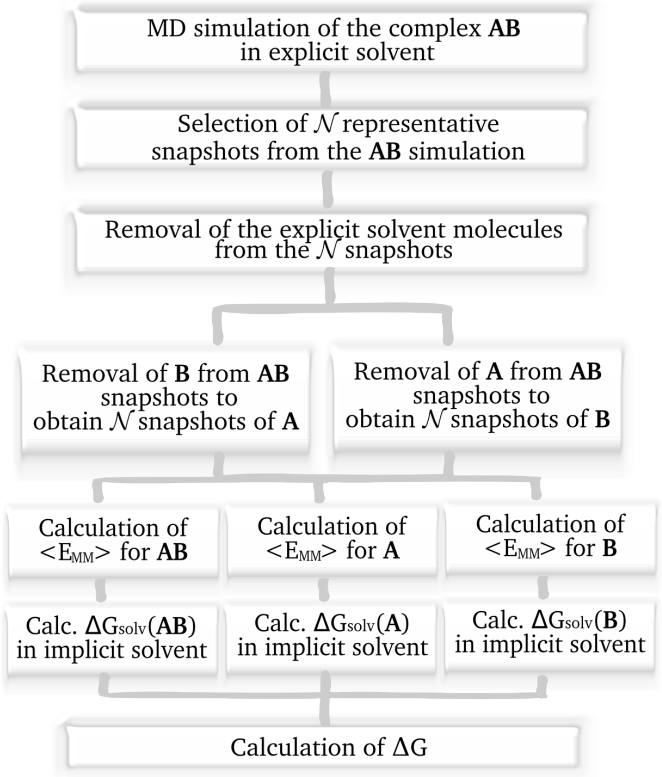

Figure 2.3: MM-GBSA calculation of the free energy of binding

where $\Delta G_{mut,X}$ is the free energy of binding when X is mutated and ΔG_{wild} is the binding energy of the native protein. Since alanine substitution loosely corresponds to a side-chain removal(‡) it can evaluate the contribution of side-chain functional groups at specific positions(§).
Hot-spots generally cause an increase of the ΔG (a positive $\Delta\Delta G$) of at least 2.0 kcal/mol, corresponding to three order of magnitudes in the binding affinity constant[56].

Experimentally, AS is generally slow and labor-intensive since it requires DNA engineering, protein expression and subsequent purification. Hence, AS is generally applied to map regions that have already been pre-selected; "blind" AS on full surfaces for explorative purposes is generally unfeasible. However, the same technique can be applied "in silico", taking the name of Computational Alanine Scanning (CAS)[57],

(‡) the mutated residue retains only the C_β, since alanine has a methyl group attached to the C_{alpha})
(§) Glycine is theoretically suitable for this purpose being the smallest residue and having no side-chain (just a H atom), but it can introduce conformational flexibility into the protein backbone and therefore is not commonly used.

obtaining reliable results in a much less expensive way, both for what concerns time and resources.

Different protocols exist for CAS and many efforts have been devoted to the development of a fast and accurate method to predict hot-spots[58–60]. Generally, the different protocols differ for the simulation conditions, e.g. for the use of one or multiple trajectories for wild-type and mutated proteins or for the use of an implicit rather than an explicit solvation. Considering the number of MD simulations, three different approaches have been proposed:

single-trajectory CAS Only a simulation of the wild-type protein complex is performed; the structure are then subjected to "post-processing" to obtain the mutated complexes. Monomers are also obtained from the complex MD by removal of the other partner protein from the complex structure, as reported for MM-GBSA calculation.

double-trajectory CAS Two different simulations are performed for the native protein complex and the mutated one; monomers are obtained by "post-processing".

quadruple-trajectory CAS Four MDs are performed for the native complex and the native monomer to be mutated and for the mutated complex and the mutated monomer. The monomer that is not mutated is crossed out by a thermodynamic cycle.

Recently[56] it has been demonstrated that, because of a convenient error cancellation, the use of the ***single-trajectory CAS*** generally produces better results, in comparison with multiple trajectories approaches, where the error due to insufficient sampling strongly affects the results. Moreover, the use of a single trajectory allows to neglect entropic terms during $\Delta\Delta G$ calculations because of the identity of the structures, leading to mutual deletion of these contributions.
Concerning implicit/explicit solvation, there is, unfortunately, little agreement on which approach would achieve better results. While a better description of the solvent is commonly preferred, implicit solvent supporters claim that, since the solvation term ΔG_{solv} in MM-GBSA in calculated using implicit solvent, it is more coherent

to use the same method for the MD simulation.[§] Since the largest deviation from experimental results are generally reported for charged residues, Moreira et al. [56] have proposed the use of a different internal dielectric constant for different residues in the MM-GBSA calculation.[‖] While better results are obtained using a variable ϵ, we decided to devise a more general solution, consisting in the use of a physiological saline concentration (0.15 M) during free energy calculations. The correction obtained with the latter approach are consistent with those obtained with a variable ϵ.

[§] Both approaches will be used (and compared) in the present work.
[‖] Specifically, $\epsilon = 2$ for non-polar residues, $\epsilon = 3$ for polar ones and $\epsilon = 4$ for charged amino acids.

CHAPTER 3

Interactions in a Tubulin Tetramer

As anticipated, one possible approach to affect tubulin polymerisation and microtubules formation is to interfere with the PPIs responsible for MT self assembly. In particular, our strategy involves the identification of key residues (hot-spots) at the protein-protein contact surface and the development of antimitotic peptides from "hot-stretches", tubulin interfacial sub-sequences rich of hot-spots (a schematic description is reported in Figure 1.7 on page 17).

3.1 Molecular Dynamics Simulations: Tetramer

In order to do study the PPIs involved in MTs formation, a simulation of a model system, a tetramer comprising two tubulin dimers aligned along the longitudinal axis (Figure 3.1) as in natural tubulin protofilaments, was performed. Different simulations were carried out varying the most relevant parameters. In particular, MD simulations were carried out for 2 ns at first and extended later to 5 ns, two different box sizes (with volume $0.8 \cdot 10^3$ nm^3 and $1.6 \cdot 10^3$ nm^3) were used alongside implicit or explicit solvation. In the end, the 5 ns MD with larger box and explicit solvent was selected as the reference simulations since these conditions were the most suitable for the study of the tetrameric system. In the following, the results for the reference simulation will be discussed; a comparison of the results with those obtained for the other simulation conditions will be reported where appropriate.

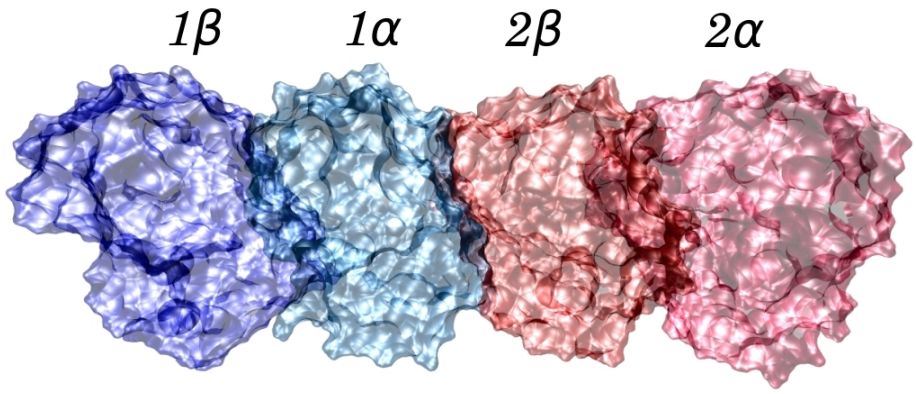

Figure 3.1: The model tetramer composed of two associated tubulin dimers.

COMPUTATIONAL DETAILS

The tetramer structure was obtained from the protein data bank (PDB id: 1TUB) containing the crystallographic structure of a tubulin dimer. The tetramer was obtained applying crystallographic transformations specified in the PDB itself and, in particular, applying a translation of 80 Å along the longitudinal axis. The obtained system was solvated with explicit TIP3P water molecules in a box of volume 1.6 $\cdot 10^3$ nm^3. Neutrality was achieved by addition of Na^+ ions. A 2 fs integration step was used alongside SHAKE algohorithm. PME with a 8 Å cutoff was used for non-bonded interactions. The system was minimized using 1000 steps of steepest descent and 9000 steps of conjugated gradient. A 200 ps NVT equilibration was performed using a Berendsen thermostat; a subsequent 200 ps NPT equilibration was performed with a Berendsen barostat. Equilibration was followed by a 5 ns NPT production phase. Simulations were carried out with the **AMBER** software package using Amber03 force-field and the Sander module.

In order to validate the structure of the tetramer, obtained applying the crystallographic transformations to the tubulin dimer, we calculated the C_α Root Mean Square Deviation (RMSD) from the initial minimized structure (Figure 3.2). From Figure 3.2 it can be easily observed that the MD in explicit solvent gave a lower overall RMSD in comparison with the implicit solvent simulation. Moreover, the initial structure clearly evolved into a stable conformation within the first 2 ns of MD, as proved by the convergence of the RMSD value, as opposed to the constant

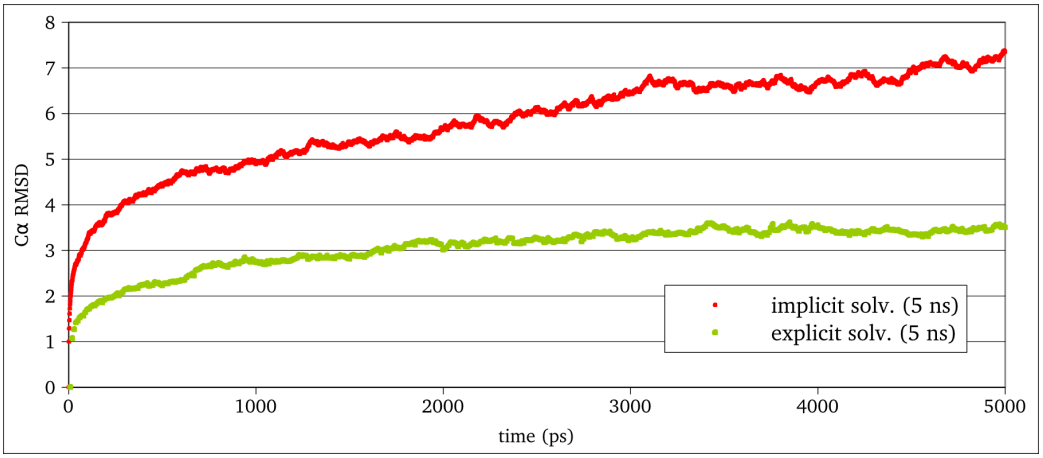

Figure 3.2: C_α-RMSD for the tubulin tetramer in implicit and explicit solvent.

RMSD increase observed for the implicit solvent MD.

3.1.2 Binding Energy

Applying the MM-GBSA approach, we calculated the binding energy for the explicit simulations; results are reported in Table 3.1 while the ΔG trend during the simulation is shown in Figure 3.3 for the two different box sizes.

	first 2 ns		last 2 ns		all 5 ns	
	$\Delta G\left(\frac{kcal}{mol}\right)$	$\sigma_{\Delta G}$	$\Delta G\left(\frac{kcal}{mol}\right)$	$\sigma_{\Delta G}$	$\Delta G\left(\frac{kcal}{mol}\right)$	$\sigma_{\Delta G}$
$V = 0.8 \cdot 10^3$ nm^3	-110,83	10,24	-93,16	7,25	-101,03	13,22
$V = 1.6 \cdot 10^3$ nm^3	-123,95	14,69	-121,45	8,33	-122,36	11,17

Table 3.1: Free energies of binding for the different simulations of the tetramer.

The ΔG converged to a defined value for both MD simulations, reporting a slightly higher value for the smaller box MD ($\sim 15\%$); the deviation is, however, of the same order of magnitude of the standard deviation ($\sim 12\%$).

The reported values don't include the configurational entropy ΔG_{config} (discussed in section 2.2.2 on page 29) since the size of the system doesn't allow for normal mode analysis. An attempt was performed considering only the interfacial region between units 1α and 2β that, theoretically, should give the most important contribution to the overall change in entropy. Including the calculated $-T\Delta S_{config}$ factor the ΔG value is decreased to $-61.87 \pm 19.53 \frac{kcal}{mol}$. The large error on the entropic term affected

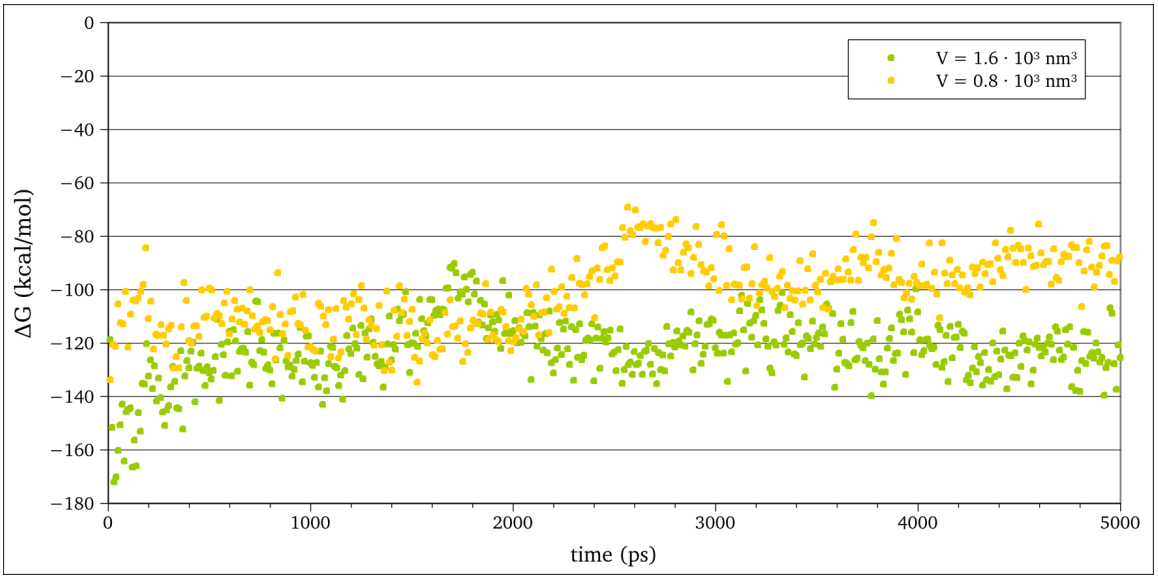

Figure 3.3: Binding free energy for the tubulin tetramer. The ΔG appears more stable during the simulation with the larger 1.6 $\cdot 10^3$ nm^3 box.

the final standard deviation that raises from $\sim$ 12% to more than 30%.

Since our purpose was to study the structure of the protein-protein interface between tubulin dimers and to map the interactions stabilizing microtubules, we needed to discriminate between interfacial and non-interfacial residues. The residues belonging to the protein-protein interface between the 1α and the 2β subunits (see Figure 3.1) of the two tubulin dimers composing our model were selected as the ensemble of residues with a non-zero difference in solvent accessible surface area in the tetrameric versus the dimeric unit. The so defined interface comprises 176 residues and matches the one previously identified in papers concerning microtubules structural models based on crystallographic and cryo-electron microscopy investigations[3].

3.1.3 Computational Alanine Scanning

Computational alanine scanning was performed on each interfacial residue, following the procedure described in the previous chapter (**Computational Methods I**). A representation of the most important residues is reported in Figure 3.4 where amino acids at the interface are classified according to the difference in the binding free energy ($\Delta\Delta$G) between the dimers upon each point mutation as **hot-spots** ($\Delta\Delta G >$ 4 kcal/mol), **warm-spots** (2 kcal/mol $< \Delta\Delta G <$4 kcal/mol) or **cold-spots** ($\Delta\Delta G$ $<$ 2 kcal/mol).

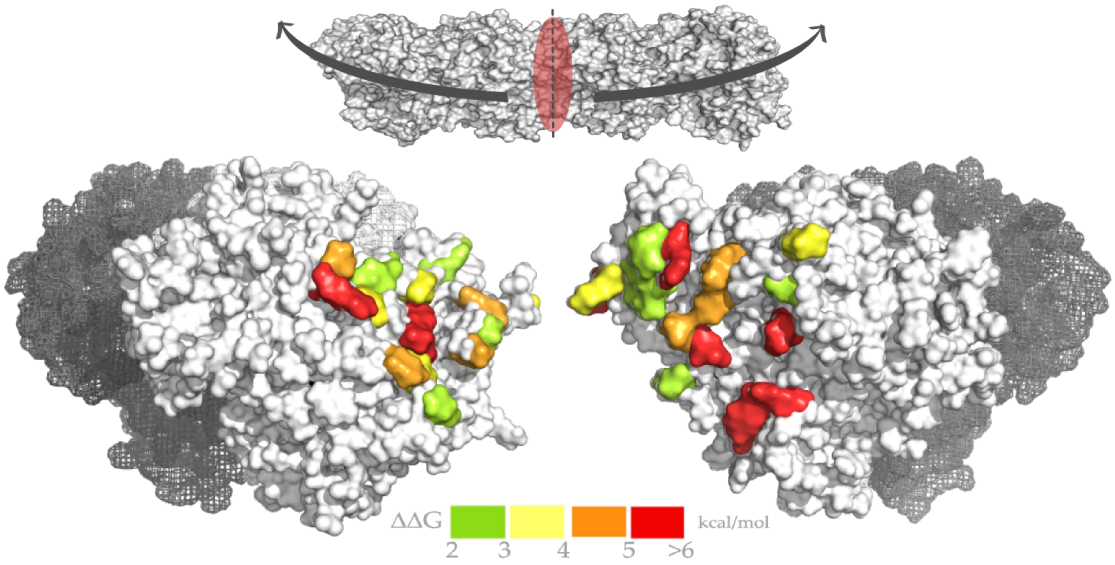

Figure 3.4: The most important residues at the protein-protein interface, colored according to their $\Delta\Delta G$ value. Only a small number of residues (15%) are involved in PPI.

As expected, not all the protein-protein interface interacts homogeneously, but only some critical amino acids. The $\Delta\Delta G$ values for al "hot-spots" and "warm-spots" identified are reported in Table 3.2.

These residues showed a tendency to be grouped in small clusters close to one another in the 3D structure and often also in the amino acidic sequence, conveniently forming "hot stretches" of residues of key importance for tubulin aggregation, suitable for our purpose, i.e. the development of peptides. Hot- and warm-spots were evenly distributed among the 1α subunit, which contained 7 hot- and 9 warm-spots, and the 2β subunit, which contained 9 hot and 7 warm-spots (their localization at the interface is reported in Table 3.2 and Figure 3.6). A deeper insight into the network of interactions between the residues at the interface that contribute to microtubule stability can be achieved through an analysis of the environment of the hot- and warm-spots during the simulation. The two most prominent hot-spots on the 1α and 2β subunits, Lys_{352} and $Asp_{177'}$ [(†)], respectively, are involved in a salt bridge throughout the simulation (Figure 3.5b). The $Asp_{177'}$ side chain also interacts, by means of a stable hydrogen bond, with the Val_{353} backbone nitrogen (Figure 3.5b). $Asp_{177'}$ belongs to the T5 loop (according to the tubulin secondary structure classification

(†) residues on the 2β subunit are labelled with a prime on their sequence number

proposed by Lowe et al.[61], see Figure 3.6) while Lys_{352} and Val_{353} are part of the S9 strand. Interestingly, these regions were proposed to interact with each other on a geometrical basis in microtubule models[61]. Another cluster of residues interacting through hydrogen bonds consists of Asn_{258} and Thr_{349} on the 1α subunit along with $Asn_{99'}$, $Val_{175'}$ and $Thr_{178'}$ on the 2β subunit (Figure 3.5a).
The observed interactions were confirmed by pairwise decomposition of the binding free energy (Table 3.3).

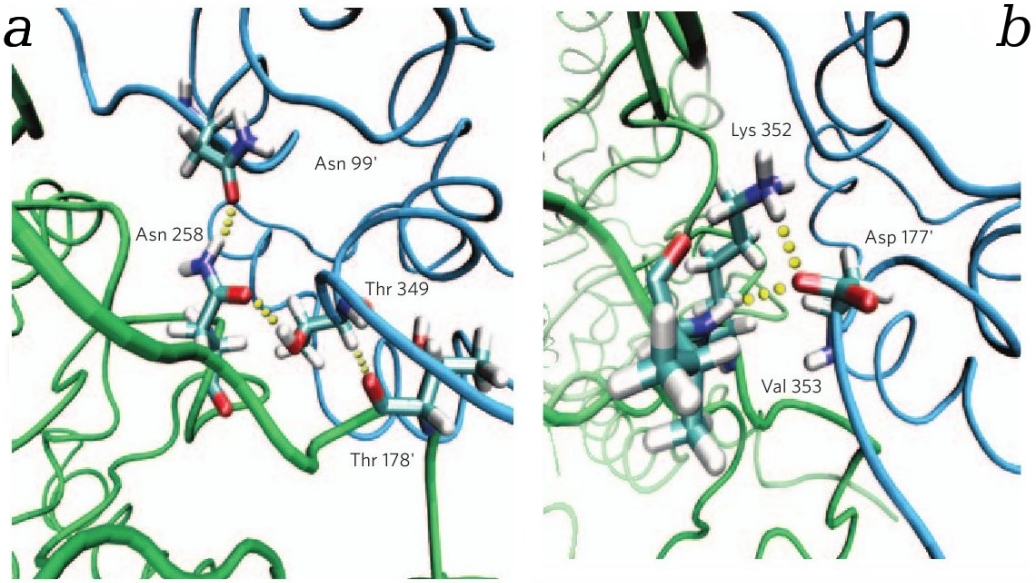

Figure 3.5: Interactions at the protein-protein interface: (a) an hydrogen bond network involving Asn_{258}, Thr_{349}, $Asn_{99'}$, $Val_{175'}$ and $Thr_{178'}$ (b) a strong salt bridge between Lys_{352} and $Asp_{177'}$.

Drawing on these observations on the energy and geometry of the protein-protein interface, we identified three "hot-stretches" containing at least five hot- and/or warm-spots each. Two of them, named group ***F*** and group ***H***, are located on the 1α subunit and are composed of residues ranging from Leu_{248} to Leu_{259} and from Trp_{346} to Val_{353}. They correspond to the T7-H8 and S9 tubulin segments respectively. The third, group ***X***, comprising residues from $Phe_{389'}$ to $Gly_{400'}$, corresponds to the H11′ region and is located on the 2β subunit (Figure 3.7). It's worth noting that the α tubulin H8 and S9 segments were reported to possibly interact with the β tubulin region between H11 and H12[3], subsequently named H11′[61].

α unit			$\Delta\Delta G\,(\frac{kcal}{mol})$	$\sigma_{\Delta\Delta G}$	β unit			$\Delta\Delta G\,(\frac{kcal}{mol})$	$\sigma_{\Delta\Delta G}$
Arg_{2}			8.70	3.27	$Gln_{11'}$			5.50	2.31
Gln_{133}		S4	4.62	2.05	$Glu_{69'}$			2.24	1.05
Leu_{248}	F	T7	4.34	0.91	$Gln_{94'}$			3.29	2.44
Asp_{251}	F	T7	3.28	2.17	$Asn_{99'}$			4.81	1.34
Thr_{253}	F	H8	2.20	2.01	$Lys_{103'}$		H3'	4.64	2.11
Gln_{256}	F	H8	2.76	2.00	$Val_{175'}$		T5	2.47	1.38
Thr_{257}	F	H8	3.63	1.69	$Asp_{177'}$		T5	15.23	6.96
Tyr_{258}	F	H8	5.00	1.42	$Thr_{178'}$		T5	4.55	1.66
Tyr_{262}			2.54	0.70	$Tyr_{208'}$			5.17	1.00
Asn_{329}		H10	2.63	1.48	$Tyr_{222'}$			5.59	1.11
Trp_{346}	H		4.49	1.18	$Met_{388'}$			2.80	0.62
Pro_{348}	H		2.31	0.72	$Arg_{391'}$	X		9.39	3.77
Thr_{349}	H		4.60	1.34	$Lys_{392'}$	X		3.86	1.97
Lys_{352}	H	S9	13.57	2.51	$Phe_{394'}$	X	H11'	2.11	0.51
Val_{353}	H	S9	3.33	2.93	$His_{396'}$	X	H11'	2.74	1.24
Asp_{438}			3.25	3.03	$Trp_{397'}$	X	H11'	6.65	1.94

Table 3.2: Hot- and Warm-spots identified on tubulin dimer-dimer interface with CAS. The three identified "hot-stretches" (F, H and X) are reported. The interacting regions proposed in structural models (S4, T7, etc.) are also reported.

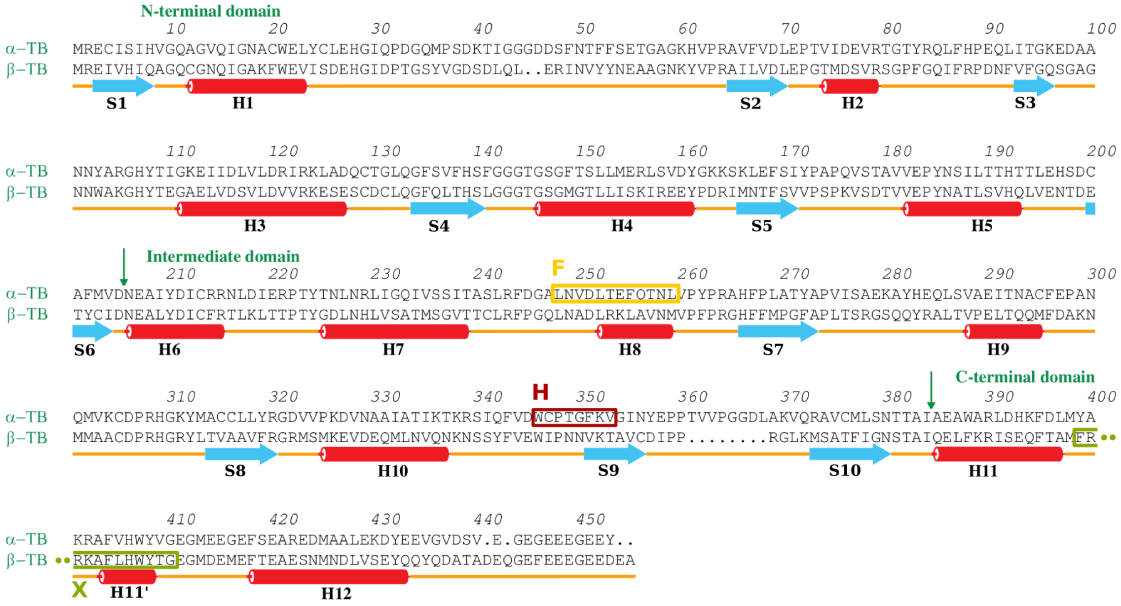

Figure 3.6: Tubulin sequence with secondary structure as published alongside crystallographic 3D structure (PDB id: 1TUB, used for the MD simulations) with the three "hot-stretches" (F, H and X) highlighted. Note that, since β unit has a shorter sequence (residues 360-367 of α-tubulin are not present) the numbering is shifted for the β unit in the sequence alignment. Helix H11' was observed only in a more recent crystallographic structure (PDB id: 1JFF)[61], nevertheless this region quickly assumes an helical structure in our MDs, even if not present in the starting 1TUB structure.

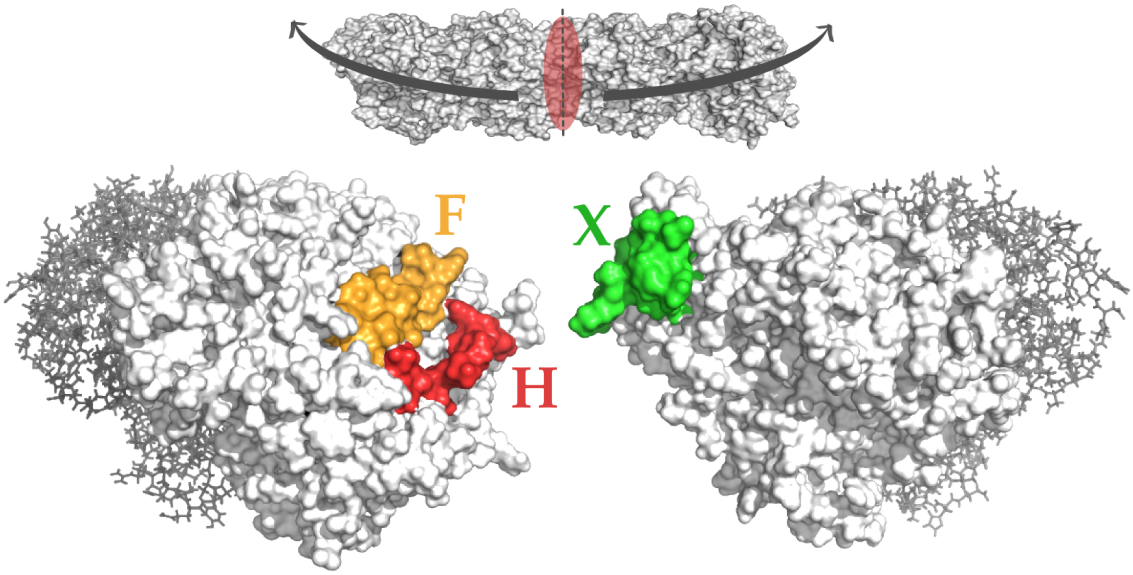

Figure 3.7: The three hot regions at the protein-protein interface.

res 1α	res 2β	E_{int} $(\frac{kcal}{mol})$	res 1α	res 2β	E_{int} $(\frac{kcal}{mol})$
Thr_{253}	$Ala_{97'}$	-1.39	Pro_{261}	$His_{396'}$	-0.75
Thr_{253}	$Gly_{98'}$	-0.78	Pro_{261}	$Trp_{397'}$	-1.15
Thr_{253}	$Lys_{103'}$	-3.42	Tyr_{262}	$Lys_{392'}$	-1.19
Glu_{254}	$Gly_{98'}$	-1.47	Tyr_{262}	$Ala_{393'}$	-1.76
Glu_{254}	$Asn_{99'}$	-3.15	Tyr_{262}	$His_{396'}$	-2.02
Glu_{254}	$Lys_{103'}$	-0.78	Tyr_{262}	$Trp_{397'}$	-1.08
Gln_{256}	$Lys_{103'}$	-3.26	Pro_{263}	$His_{396'}$	-1.49
Gln_{256}	$Trp_{397'}$	-1.14	Pro_{263}	$Trp_{397'}$	-1.01
Thr_{257}	$Asn_{99'}$	-0.95	Trp_{346}	$Met_{388'}$	-1.11
Thr_{257}	$Asn_{100'}$	-1.72	Trp_{346}	$Lys_{392'}$	-4.24
Thr_{257}	$Lys_{103'}$	-0.81	Met_{348}	$Met_{388'}$	-1.42
Thr_{257}	$Thr_{178'}$	-0.58	Met_{348}	$Thr_{178'}$	-0.78
Thr_{257}	$Val_{179'}$	-0.55	Thr_{349}	$Val_{175'}$	-2.30
Thr_{257}	$Ala_{394'}$	-0.50	Thr_{349}	$Ser_{176'}$	-3.88
Thr_{257}	$Trp_{397'}$	-1.41	Thr_{349}	$Thr_{178'}$	-3.66
Asn_{258}	$Asn_{99'}$	-1.54	Gly_{350}	$Thr_{178'}$	-1.84
Asn_{258}	$Asp_{177'}$	-0.50	Phe_{351}	$Val_{175'}$	-0.84
Asn_{258}	$Thr_{178'}$	-3.37	Phe_{351}	$Ser_{176'}$	-2.84
Asn_{258}	$Val_{179'}$	-1.22	Phe_{351}	$Asp_{177'}$	-1.38
Val_{260}	$Trp_{397'}$	-1.04	Phe_{351}	$Thr_{178'}$	-0.67
Pro_{261}	$Met_{388'}$	-0.86	Lys_{352}	$Asp_{177'}$	-10.93
Pro_{261}	$Lys_{392'}$	-1.10	Lys_{352}	$Thr_{178'}$	-0.66
Pro_{261}	$Trp_{394'}$	-1.38	Val_{353}	$Asp_{177'}$	-4.58

Table 3.3: Pairwise decomposition of the binding free energy. The strong interaction between Lys_{352} and $Asp_{177'}$ is confirmed, together with the interaction with Val_{353}.

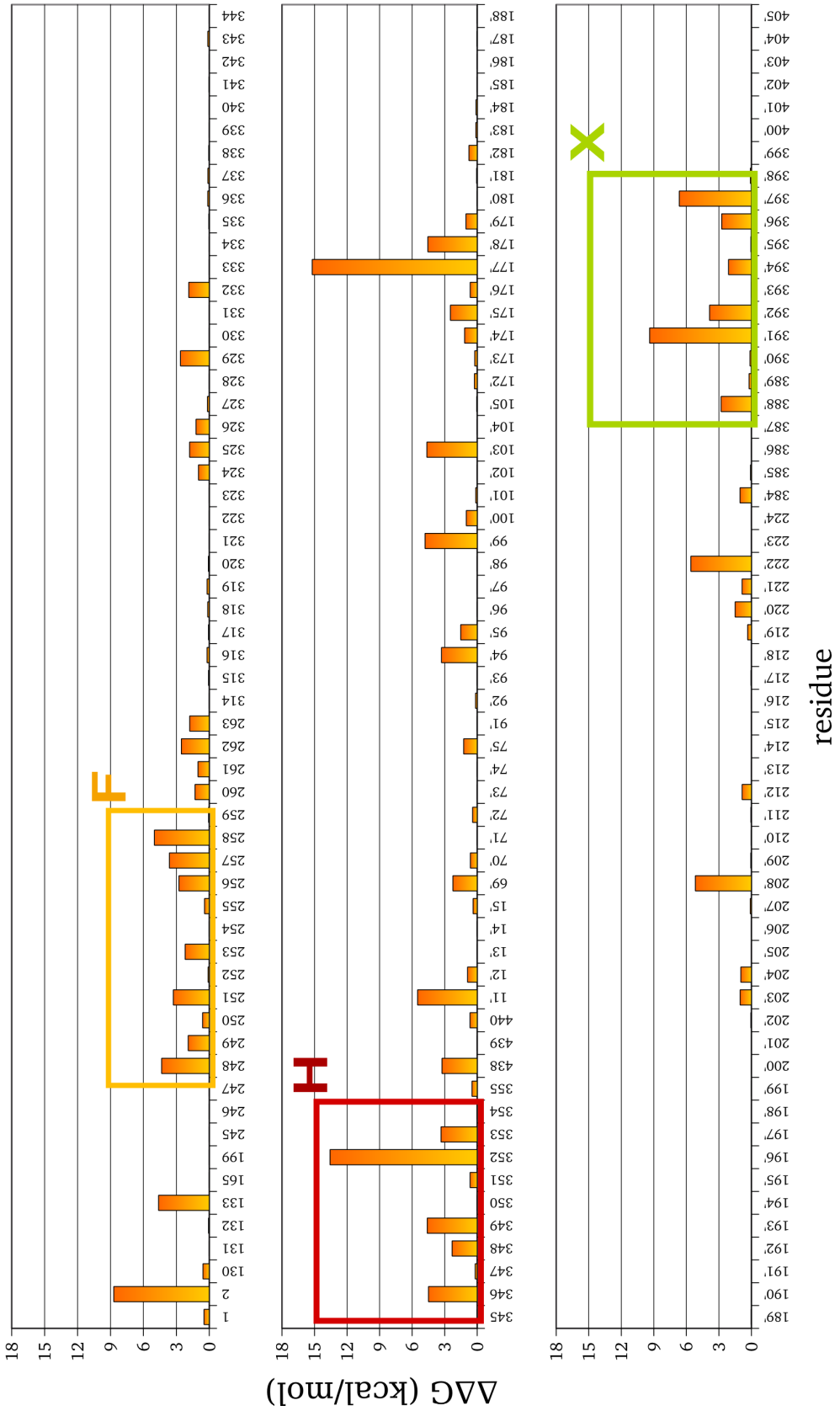

Figure 3.8: CAS results for all the interfacial residues. The three groups (F, H and X) are the regions with at least 5 hot-/warm-spots close in the sequence.

3.1.4 Validation of the Results

Phylogenetic analysis

To confirm the reliability of the hot/warm-spots identification we cross-checked $\Delta\Delta G$ values with evolutionary data on residues conservation. Since hot/warm-spots are important for protein-protein association, we expect them to be among the most highly conserved residues. Phylogenetic analysis of the tubulin sequence was carried out using the ConSurf server. Given a 3D structure, the server automatically searches for homologous sequences using PSI-BLAST. The sequence is obtained from the SEQRES record in the supplied PDB file or from the ATOM record if the latter is not present. The sequences are then aligned using MUSCLE, and a conservation score is calculated using an empirical Bayesian method (or Maximum Likelihood method). The conservation scores are normalized in order to have an average score of zero and a standard deviation of one. The lowest scores represent the most conserved site in the protein sequence.

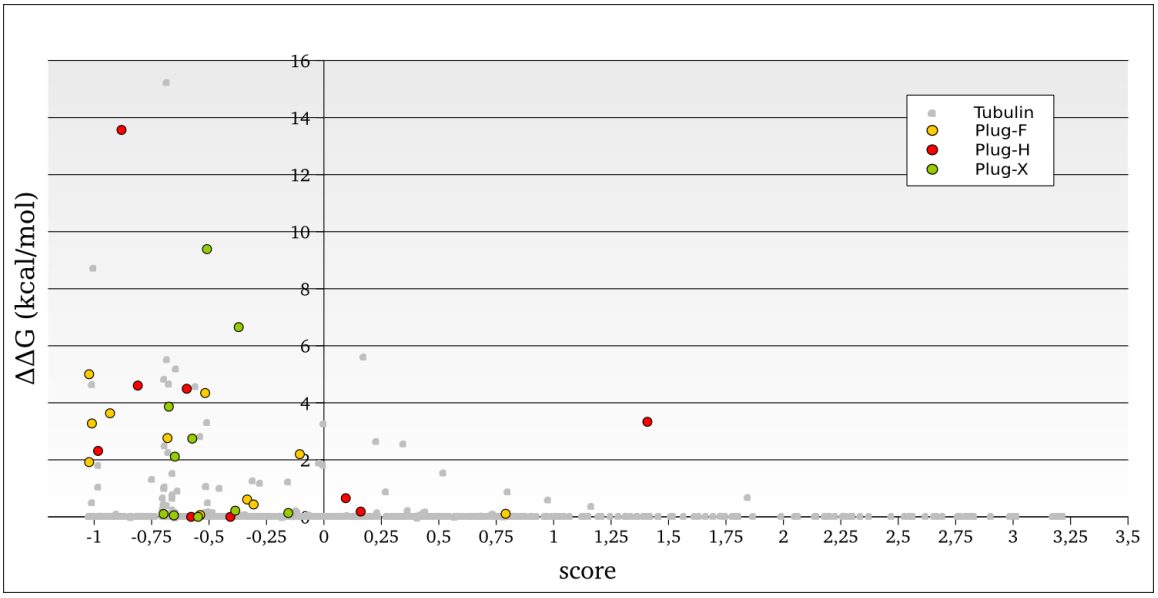

Figure 3.9: Phylogenetic analysis of the tubulin sequence using the ConSurf server. The highly conserved residues belonging to group F, H and X are highlighted with different colors.

For our analysis, 150 tubulin sequences from different organisms were collected and aligned. The $\Delta\Delta$G values obtained from the explicit solvent simulation for all the interfacial residues were plotted against the evaluated score function. Of these, 15

out of 16 hot and 13 out of 16 warm spots exhibited conservation above average. $Tyr_{222'}$ is the only significant outlier among the hot-spots. However, analyses of the mutations revealed that this residue is often replaced by a phenylalanine, confirming the importance of having an aromatic hydrophobic amino acid in this position. The three warm spots that are less conserved than average all have conservative mutations among the different species. In particular, Tyr_{262} was most often mutated to phenylalanine, Asn_{329} was mutated, with diminishing probability, to aspartate, glutamate, histidine or other polar residues, and finally, Val_{353}, the most evident outlier, was most often mutated to alanine, cysteine, isoleucine, leucine or another hydrophobic residue. The results obtained using the $\Delta\Delta$G values from implicit solvent simulations show no significant difference.

Dependence on CAS parameters

Since different protocols are available for CAS, the analysis was repeated varying the main parameters of the scanning. First of all we performed CAS on the implicit solvent MD trajectory (Table 3.4).

unit α	Explicit Solvent		Implicit Solvent		unit β	Explicit Solvent		Implicit Solvent	
residue	$\Delta\Delta G(\frac{kcal}{mol})$	$\sigma_{\Delta\Delta G}$	$\Delta\Delta G(\frac{kcal}{mol})$	$\sigma_{\Delta\Delta G}$	residue	$\Delta\Delta G(\frac{kcal}{mol})$	$\sigma_{\Delta\Delta G}$	$\Delta\Delta G(\frac{kcal}{mol})$	$\sigma_{\Delta\Delta G}$
Arg_{2}	8.70	3.27	15.87	5.46	$Gln_{11'}$	5.50	2.31	9.77	2.47
Gln_{133}	4.62	2.05	9.52	4.64	$Glu_{69'}$	2.24	1.05	14.04	4.05
Leu_{248} F	4.34	0.91	6.56	1.17	$Gln_{94'}$	3.29	2.44	6.51	3.58
Asp_{251} F	3.28	2.17	-0.73	2.48	$Asn_{99'}$	4.81	1.34	8.78	1.87
Thr_{253} F	2.20	2.01	1.83	0.98	$Lys_{103'}$	4.64	2.11	3.35	0.43
Gln_{256} F	2.76	2.00	-0.20	0.58	$Val_{175'}$	2.47	1.38	1.67	0.57
Thr_{257} F	3.63	1.69	3.48	1.17	$Asp_{177'}$	15.23	6.96	12.85	6.83
Tyr_{258} F	5.00	1.42	10.56	3.63	$Thr_{178'}$	4.55	1.66	8.53	1.42
Tyr_{262}	2.54	0.70	0.86	0.37	$Tyr_{208'}$	5.17	1.00	2.58	0.98
Asn_{329}	2.63	1.48	2.69	1.20	$Tyr_{222'}$	5.59	1.11	6.15	1.55
Trp_{346} H	4.49	1.18	5.88	1.34	$Met_{388'}$	2.80	0.62	2.55	0.95
Pro_{348} H	2.31	0.72	4.81	0.78	$Arg_{391'}$ X	9.39	3.77	15.42	3.89
Thr_{349} H	4.60	1.34	4.50	1.68	$Lys_{392'}$ X	3.86	1.97	17.70	3.53
Lys_{352} H	13.57	2.51	35.07	2.88	$Phe_{394'}$ X	2.11	0.51	5.29	1.29
Val_{353} H	3.33	2.93	2.84	1.45	$His_{396'}$ X	2.74	1.24	0.11	0.22
Asp_{438}	3.25	3.03	5.58	3.18	$Trp_{397'}$ X	6.65	1.94	7.95	1.28

Table 3.4: Comparison of CAS results using explicit or implicit solvent in the MD simulation.

The results, reported in Table 3.4, highlighted how the use of an implicit solvent magnifies the $\Delta\Delta G$ values of positively charged residues, as Lys_{352} whose mutation causes a $\sim$30% destabilization (see ΔG on page 36) of the complex. The $\Delta\Delta G$

of negatively charged residues or polar residues are generally strongly affected too. Despite giving significantly different results, the identification of the highly interacting regions is not affected by the change in solvent description. In particular, the two hot-spots of the first group (F) reported higher values using the implicit solvent, while the loss of two warm-spots (Asp_{251} and Gln_{256}) didn't affect the definition of group F, since these residues are surrounded by other hot-/warm-spots. The implicit solvent values for group H are all higher than the explicit solvent ones or very slightly decreased and, with the exception of $His_{396'}$, also the values for group X are confirmed. In the whole, as anticipated, despite the increase of some hot-/warm-spots and the loss of others, the subsequences selected as "hot-regions" were substantially the same confirming that, while relevant for the single values, the change of solvent description doesn't modify the final outcome, important for the identification of possible peptides. All CAS analyses were performed using a 0.15 M saline concentration, in order to slightly correct the values for polar and charged residues, whose deviation is generally larger from experiment. This approach is similar to the multi-ϵ approach discussed in the previous chapter (**Computational Methods I**). In Table 3.5 the $\Delta\Delta G$ values in pure water and the saline solution are reported.

unit α	[NaCl] = 0.15 M		Pure Water		unit β	[NaCl] = 0.15 M		Pure Water	
residue	$\Delta\Delta G(\frac{kcal}{mol})$	$\sigma_{\Delta\Delta G}$	$\Delta\Delta G(\frac{kcal}{mol})$	$\sigma_{\Delta\Delta G}$	residue	$\Delta\Delta G(\frac{kcal}{mol})$	$\sigma_{\Delta\Delta G}$	$\Delta\Delta G(\frac{kcal}{mol})$	$\sigma_{\Delta\Delta G}$
Arg_{2}	8.70	3.27	13.35	3.35	$Gln_{11'}$	5.50	2.31	9.77	2.47
Gln_{133}	4.62	2.05	4.68	2.06	$Glu_{69'}$	2.24	1.05	14.04	4.05
Leu_{248} F	4.34	0.91	4.34	0.91	$Gln_{94'}$	3.29	2.44	6.51	3.58
Asp_{251} F	3.28	2.17	-1.14	2.06	$Asn_{99'}$	4.81	1.34	4.77	1.34
Thr_{253} F	2.20	2.01	2.19	1.99	$Lys_{103'}$	4.64	2.11	8.50	2.11
Gln_{256} F	2.76	2.00	2.70	2.00	$Val_{175'}$	2.47	1.38	2.47	1.38
Thr_{257} F	3.63	1.69	3.61	1.69	$Asp_{177'}$	15.23	6.96	11.86	6.92
Tyr_{258} F	5.00	1.42	4.96	1.42	$Thr_{178'}$	4.55	1.66	4.54	1.66
Tyr_{262}	2.54	0.70	2.52	0.70	$Tyr_{208'}$	5.17	1.00	5.16	1.00
Asn_{329}	2.63	1.48	2.57	1.48	$Tyr_{222'}$	5.59	1.11	5.57	1.12
Trp_{346} H	4.49	1.18	4.49	1.18	$Met_{388'}$	2.80	0.62	2.79	0.62
Pro_{348} H	2.31	0.72	2.29	0.72	$Arg_{391'}$ X	9.39	3.77	13.20	3.80
Thr_{349} H	4.60	1.34	4.62	1.34	$Lys_{392'}$ X	3.86	1.97	8.10	1.95
Lys_{352} H	13.57	2.51	18.22	2.52	$Phe_{394'}$ X	2.11	0.51	2.11	0.51
Val_{353} H	3.33	2.93	3.38	2.94	$His_{396'}$ X	2.74	1.24	2.75	1.23
Asp_{438}	3.25	3.03	-0.13	3.00	$Trp_{397'}$ X	6.65	1.94	6.62	1.94

Table 3.5: Comparison of CAS results using a physiological saline concentration (0.15 M) or pure water.

The comparison between the results for the two solvents (pure water and NaCl 0.15 M) reported in Table 3.5 showed how apolar residues values remained substantially the

same in the two conditions, consistent with the fact that corrections are needed only for polar and charged residues (for which an increase of ϵ is proposed [56]). Once more, even if different values are obtained, especially for charged residues, the selection of the highly interacting regions was not affected by the change of solvent. In the end, the definition of groups, F, H and X, selected as "hot-stretches" and suitable for the development of peptides is solid and unambiguous, being independent of CAS parameters and consistent with phylogenetic data.

CHAPTER 4

Development of Antitumoral Peptides

We developed three peptides from the three "hot-regions" identified, removing the corresponding subsequences from the entire dimer sequence. They were named **Plug-F**, **Plug-H** and **Plug-X**, matching the letters used for the three groups of residues.

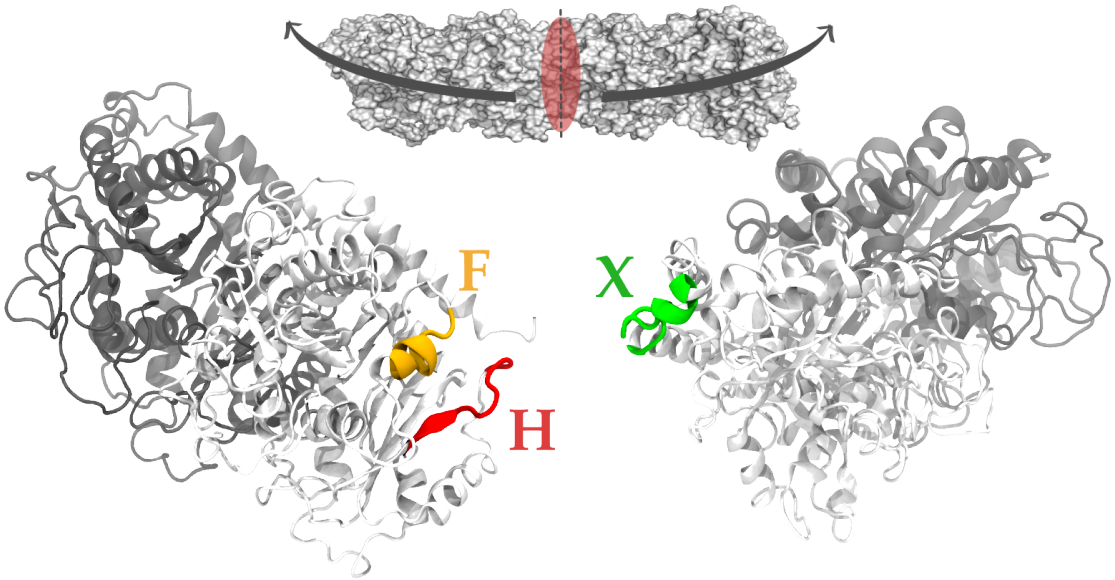

Figure 4.1: Location of the three peptides obtained from the three hot-regions at the protein-protein interface.

We inferred that these peptides could interfere with tubulin polymerization via competitive binding to the microtubule plus end and/or to the isolated dimers. These peptides correspond to tubulin subsequences, with the only exception of Cys_{347} in Plug-H that was mutated to alanine because of its negative alanine scanning and to

avoid the insertion of a bulky protecting group unrelated to the protein structure. An additional alanine residue was added before Trp_{346} for synthetic convenience.

4.1 Molecular Dynamics Simulations: Peptides

The binding capability of an isolated peptide is not guaranteed in principle, even if it contains several hot spots, because it may undergo major structural rearrangements when removed from the parent protein and because of the complexity of the interactions network leading to protein aggregation. To investigate if the identified peptides retained their ability to bind to tubulin even when extracted from their protein environment, we ran three control molecular dynamics simulations of each tubulin-peptide complex. We found that in two cases, namely Plug-H and Plug-X, the structure of the peptides underwent only minor structural modifications compared to the structure of the corresponding protein segments in full length tubulin.

4.1.1 Plug-H

As mentioned before, Plug-H includes residues from Trp_{346} to Val_{353} belonging to the α subunit of tubulin, plus an alanine before Trp_{346}. The binding energy (Figure

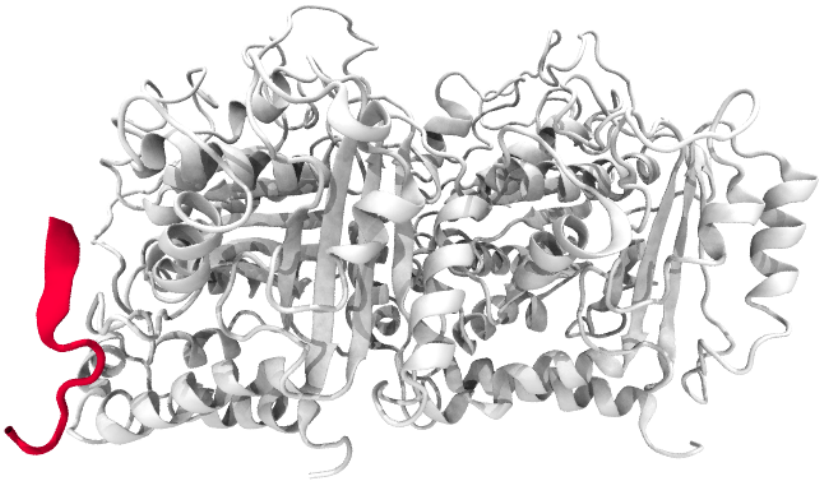

Figure 4.2: The Plug-H—tubulin complex simulated. The peptide-dimer complex was obtained removing the rest of the sequence of the $1\alpha1\beta$ dimer, to which Plug-H belongs, from the entire $1\alpha1\beta2\alpha2\beta$ tetramer sequence.

4.3) remained stable during all the MD run, with an average value of -51.45 ± 5.35 kcal/mol; averaging over the last 500 ps only we obtained a value of -47.26 ± 3.20 kcal/mol. The value proved a strong binding between Plug-H and tubulin, especially considering the value obtained for the dimer-dimer complex (-122,36 kcal/mol). Despite the stable ΔG value, analyzing the trajectory we found out that Plug-H

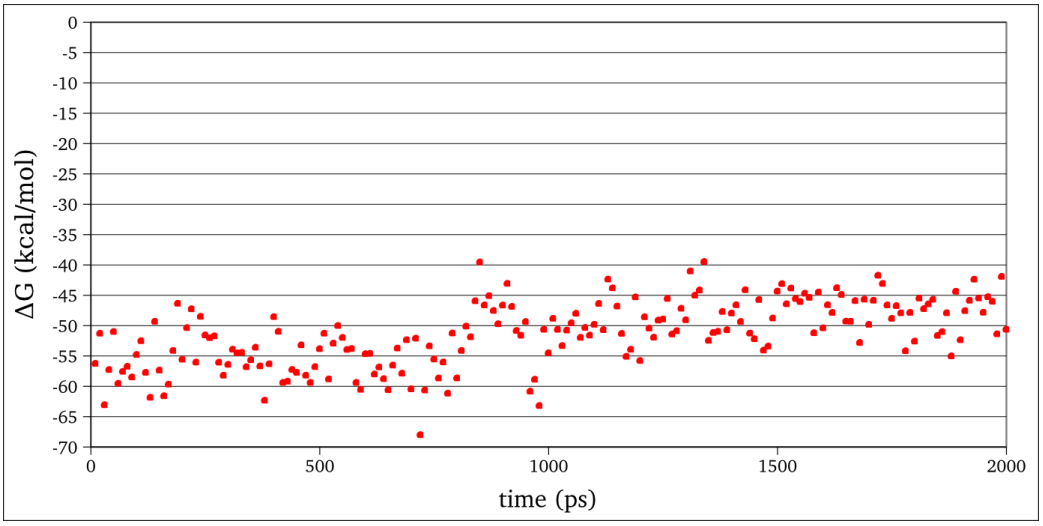

Figure 4.3: The ΔG for the Plug-H—dimer complex. The value is stable throughout the simulation and constitutes more than one third of the observed dimer-dimer binding energy.

conformation changed during the MD run. The N-terminus, in particular, bent over the peptide itself, with the consequent detachment of the hot-spot Trp_{346} from tubulin surface. On the other hand, the remaining residues conserve their close contact with the protein (Figure 4.4) and, in particular, the salt bridge between Lys_{352} and $Asp_{177'}$ (Figure 4.5) is stable during the entire simulation. CAS was performed once more, on the peptide-tubulin complex trajectory. The $\Delta\Delta G$ values (Figure 4.6) confirmed the the loss of Trp_{346} hot nature due to the aforementioned bending of the peptide N-terminus during the simulation. The $\Delta\Delta G$ value of Lys_{352} is remained unaffected, in agreement with the observed conservation of the salt bridge with $Asp_{177'}$.

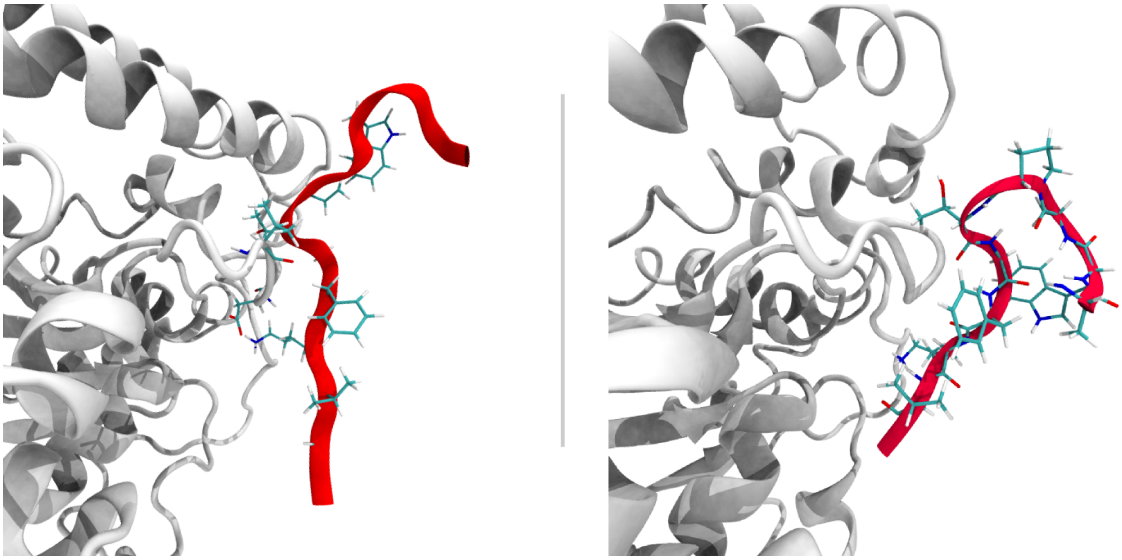

Figure 4.4: Structures for Plug-H at the beginning (left) and end (right) of the simulation. It can be observed that, during the MD, the N-terminus bent over the peptide itself, causing the loss of Trp_{346} interactions.

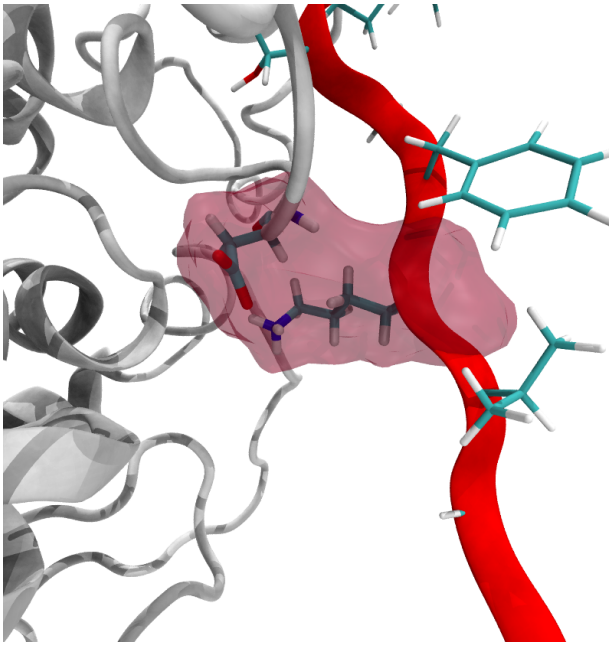

Figure 4.5: The strong salt-bridge between Lys_{352} and $Asp_{177'}$ observed throughout the entire simulation.

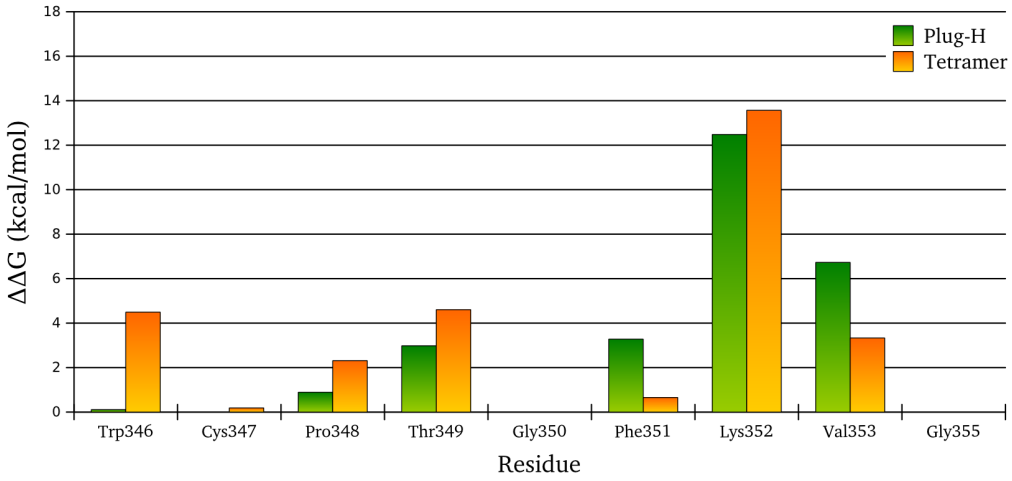

Figure 4.6: CAS for the plug-H—tubulin complex. While the $\Delta\Delta G$ values for the first residues are all decreased, due to the N-terminus bending, the last residues experienced an increase of the $\Delta\Delta G$, resulting in an overall strong binding between Plug-H and the tubulin dimer.

The strong interactions between the last residues of Plug-H (especially Lys_352 and Val_353) and tubulin surface are confirmed by the calculation of the interactions through pairwise decomposition of the binding free energy (Table 4.2).

res 1α	res 2β	E_{int} $(\frac{kcal}{mol})$	res 1α	res 2β	E_{int} $(\frac{kcal}{mol})$
Thr_{349}	$Val_{175'}$	-2,04	Phe_{351}	$Asp_{177'}$	-1,46
Thr_{349}	$Ser_{176'}$	-3,22	Lys_{352}	$Asp_{177'}$	-14,28
Thr_{349}	$Thr_{178'}$	-3,35	Val_{353}	$Asp_{177'}$	-7,38
Gly_{350}	$Thr_{178'}$	-1,32	Gly_{354}	$Tyr_{222'}$	-1,65
Phe_{351}	$Ser_{176'}$	-3,82			

Table 4.1: Pairwise decomposition of the Plug-H—tubulin ΔG. The interactions of Lys_352 and Val_353 with $Asp_{177'}$, observed in the tetramer, are enhanced in the Plug-H—tubulin complex. On the other hand, the interactions of Trp_{346} and Pro_{348} disappear, in agreement with the observed bending of the first residues of the peptide.

4.1.2 Plug-X

Plug-X comprises residues from $Phe_{389'}$ to $Gly_{400'}$, the peptide-tubulin complex is reported in Figure 4.7. The ΔG of binding for Plug-X was only slightly lower than the one observed for Plug-H, and remained stable throughout the MD simulations, with an average value of -44.15 $\pm$ 3.29 in the last 500 ps (Figure 4.8). Plug-X also showed a minor rearrangement at the N-terminus (Figure 4.9), related to the increase of $Arg_{391'}$ solvent exposure and $Phe_{389'}$ approaching closer to the protein surface. The

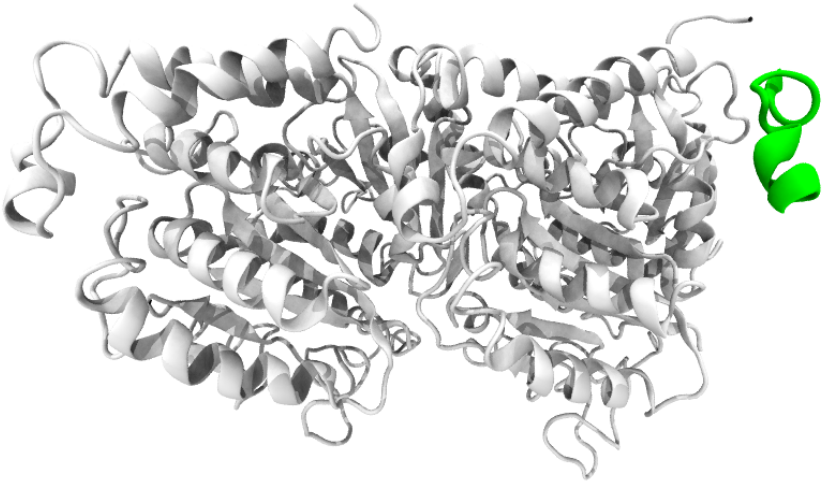

Figure 4.7: The plug-X—tubulin complex simulated. Since Plug-X belongs to the 2β unit, the peptide-dimer complex was obtained removing the rest of the sequence of the $2\alpha2\beta$ dimer from the entire $1\alpha1\beta2\alpha2\beta$ tetramer sequence.

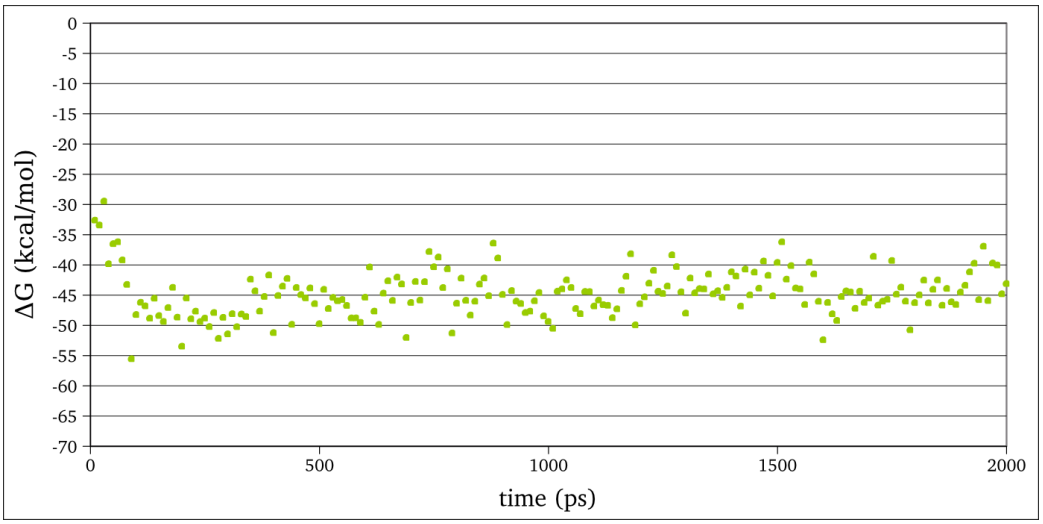

Figure 4.8: As observed for Plug-H, the binding energy was stable throughout the simulations, with a value slightly below the one reported by the other peptide.

analysis of the trajectory revealed that these modifications are due to the formation of a more compact hydrophobic core by $Trp_{397'}$,$His_{396'}$, $Phe_{394'}$ and some residues of the protein surface (Figure 4.10).

CAS analyses and pairwise ΔG decomposition confirmed the strong interaction with tubulin, especially for $Phe_{394'}$ and $Trp_{397'}$, related to the strengthening of peptide-protein hydrophobic contacts.

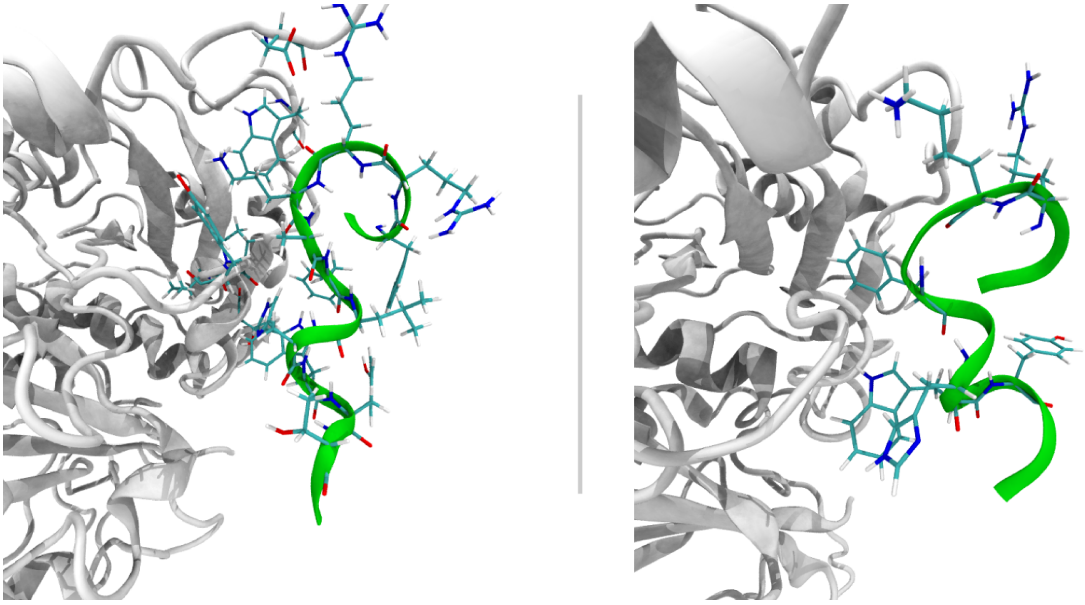

Figure 4.9: Structures for Plug-X at the beginning (left) and end (right) of the simulation. Only minor rearrangements can be observed.

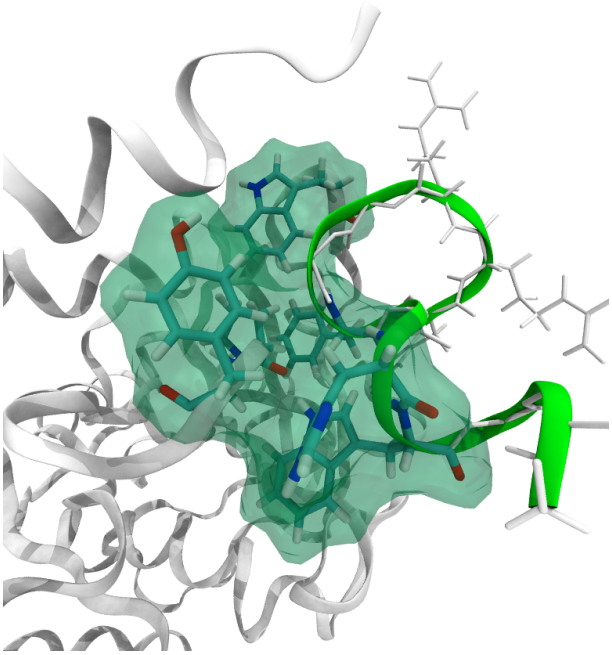

Figure 4.10: The hydrophobic core formed by Plug-X residues $Trp_{397'}$, $His_{396'}$, $Phe_{394'}$ and some interfacial residue on the tubulin dimer.

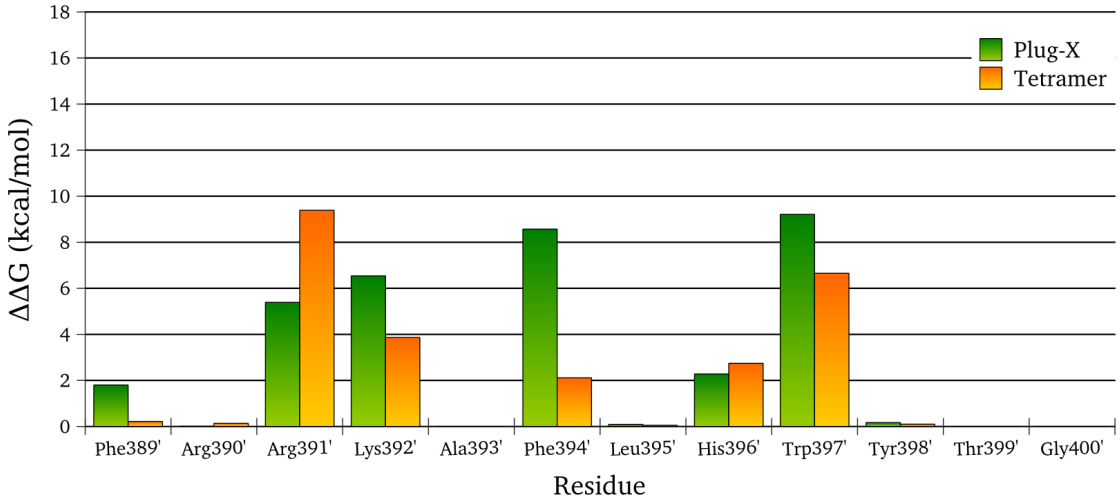

Figure 4.11: CAS for the plug-X—tubulin complex. Most residues had an increase of the $\Delta\Delta G$ value, with the exception of $Arg_{391'}$. The increased value for $Phe_{394'}$ and $Trp_{397'}$ are consistent with the observed streaghtening if the hydrophobic core involving these residues.

Finally, $Arg_{391'}$ $\Delta\Delta G$ decreased value was consistent with its greater solvent exposure upon N-terminus structural rearrangements. Interestingly, very strong interaction were observed between Plug-X and residue Trp_{346} on the 1α unit (see Table 4.2) belonging to the H group (and Plug-H).

res 1α	res 2β	E_{int} $(\frac{kcal}{mol})$	res 1α	res 2β	E_{int} $(\frac{kcal}{mol})$
$Arg_{391'}$	Trp_{346}	-2.65	$Phe_{394'}$	Trp_{346}	-0.88
$Lys_{392'}$	Trp_{346}	-4.55	$His_{396'}$	Tyr_{262}	-1.94
$Lys_{392'}$	Val_{435}	-1.66	$His_{396'}$	Pro_{263}	-1.12
$Lys_{392'}$	Asp_{438}	-4.49	$Trp_{397'}$	Gln_{256}	-1.69
$Lys_{392'}$	Ser_{439}	-0.94	$Trp_{397'}$	Thr_{257}	-3.69
$Ala_{393'}$	Pro_{261}	-1.05	$Trp_{397'}$	Val_{260}	-2.29
$Ala_{393'}$	Val_{435}	-0.76	$Trp_{397'}$	Pro_{261}	-1.09
$Phe_{394'}$	Val_{260}	-1.57	$Trp_{397'}$	Tyr_{262}	-1.16
$Phe_{394'}$	Pro_{261}	-1.05	$Trp_{397'}$	Pro_{263}	-1.24

Table 4.2: Pairwise decomposition of the Plug-X—tubulin ΔG. The interactions of $Trp_{397'}$, $His_{396'}$ and $Phe_{394'}$ confirmed the streaghtening of the hydrophobic contacts.

4.1.3 Plug-F

The third peptide, Plug-F, comprises residues from Leu_{248} to Leu_{259}.

The free energy of binding for Plug-F (Figure 4.13) showed a trend dissimilar to the

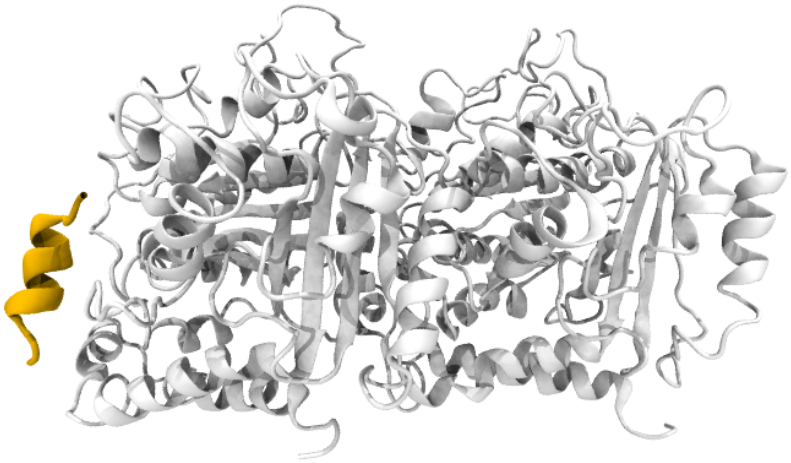

Figure 4.12: The plug-F—tubulin complex simulated. Since Plug-F, as Plug-H, belongs to the 1α unit, the peptide-dimer complex was obtained removing the rest of the sequence of the 1α1β dimer from the entire 1α1β2α2β tetramer sequence.

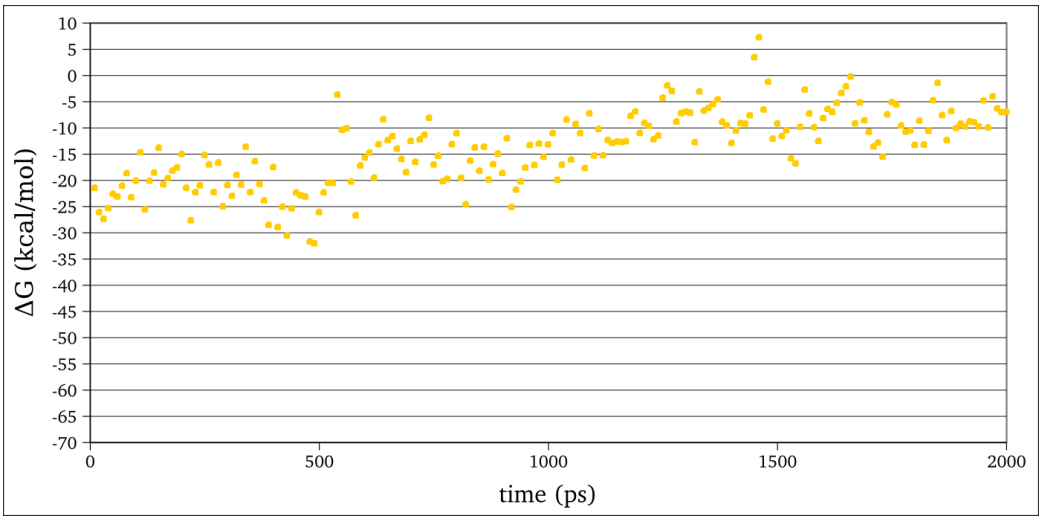

Figure 4.13: The free energy of binding for the Plug-F—tubulin complex. The constant increase lead to a progressive detachment of the peptide from the dimer surface.

one observed for the other peptides. Starting from a value around -20 kcal/mol, already considerably higher than the starting value for Plug-H and Plug-X, the ΔG increased constantly, even becoming positive in a few snapshots, and showing a progressive weakening of the Plug-F—tubulin interaction. The binding energy, indeed, reported a value close to zero, especially considering the standard deviation. The average for the last 500 ps was -6.53 $\pm$ 5.22 kcal/mol, significantly higher that the value reported for both the two other peptides. Analyzing the trajectory, Plug-F

showed major structural changes, partially loosing the helical structure observed in the tetramer and detaching from the protein.

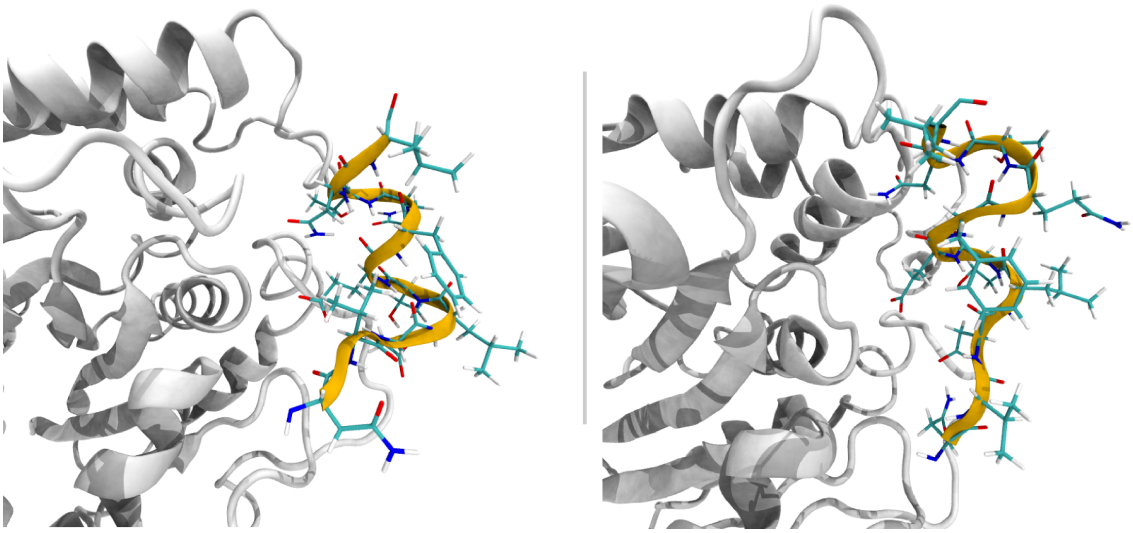

Figure 4.14: Structures for Plug-F at the beginning (left) and end (right) of the simulation. Major structural rearrangements can be observed, resulting in a significant decrease of the binding energy.

CAS analysis revealed the loss of four out of five hot- and warm-spots initially present (Figure 4.15). The limited increase in $\Delta\Delta G$ value exhibited by Asn_{258} was clearly not enough to compensate the relevant decrease of the other critical spots. In the whole, Plug-F was reputed unable to interact strongly with tubulin on its own.

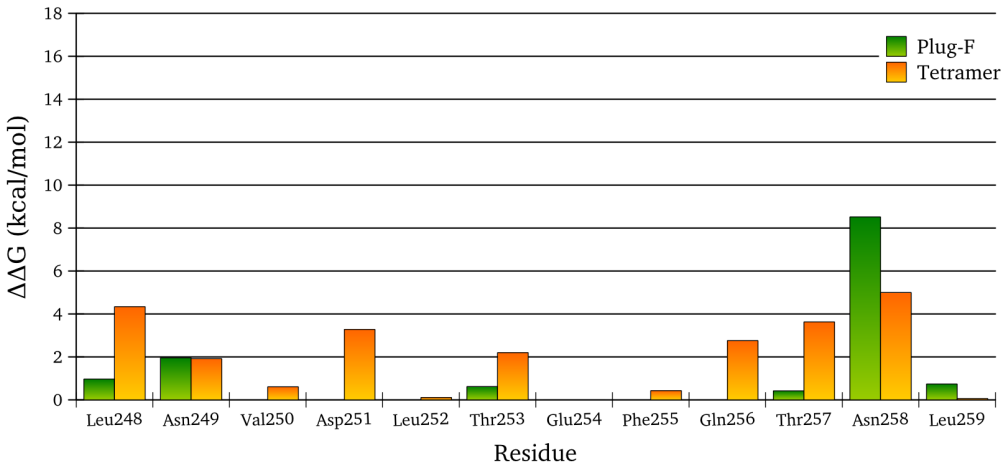

Figure 4.15: CAS for the Plug-F—tubulin complex. All the hot./warm-spots are lost with the exception of Asn_{258}.

res 1α	res 2β	E_{int} ($\frac{kcal}{mol}$)	res 1α	res 2β	E_{int} ($\frac{kcal}{mol}$)
Asn_{249}	$Gln_{11'}$	-3.71	Asn_{258}	$Asp_{177'}$	-2.23
Thr_{253}	$Asn_{99'}$	-1.19	Asn_{258}	$Thr_{178'}$	-2.56
Glu_{254}	$Asn_{99'}$	-2.86	Asn_{258}	$Val_{179'}$	-5.22

Table 4.3: Pairwise decomposition of the Plug-F—tubulin ΔG. Strong interactions are only observed for Asn_{258}, the only remaining hot-spot.

We infer from these data that either Plug-H or Plug-X could retain tubulin binding ability and therefore affect tubulin polymerization, since the hot- and warm- spots observed in the tubulin tetramer were conserved or only slightly modified in the peptides. In order to confirm computational results, the three peptides were synthesized and tested.

EXPERIMENTAL VALIDATION

Peptides Biological Tests

The following biological tests were performed by Dr. G. Cappelletti and Dr. D. Cartelli and are reported for the sake of completeness. Prof. G. Speranza and Dr. P. Francescato are also acknowledged for peptides syntheses.

The biological activity of the designed peptides was assessed in vitro using purified tubulin and cultured cells. We initially investigated the ability of peptides to interfere with microtubule assembly in vitro and found that Plug-X and Plug-H, but not Plug-F, affect tubulin polymerization (Figure 4.16).

The apparent first-order rate constant of elongation, the steady-state extent of assembly, and the critical concentration of tubulin were calculated. Elongation rate in the presence of Plug-X and Plug-H, respectively, were significantly decreased with respect to the values that were derived from control kinetics.

In contrast, the elongation rate was unaffected by Plug-F and two scrambled(†) peptides (named Plug-Xs and Plug-Hs) derived from Plug-X and Plug-H. Similarly, the presence of Plug-X and Plug-H, but not Plug-F and the two scrambled peptides, significantly affected the absorbance maximum at the end of assembly, which is proportionally related to the mass concentration of the tubulin polymer. Noticeably, the activity or inactivity of the peptides in inhibiting tubulin polymerization parallels

(†) A "scrambled" peptide is obtained changing the order of the different residues of a peptide. The testing of a scrambled peptide is a common proof of the importance of the peptide specific amino acidic sequence.

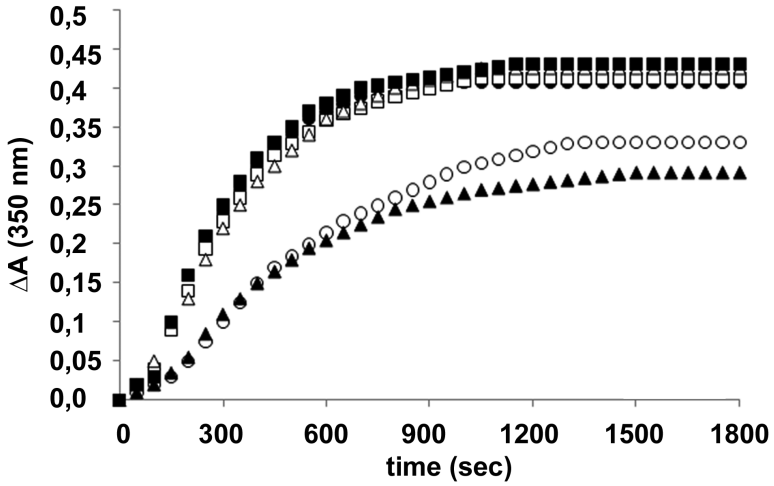

Figure 4.16: Tubulin assembly was recorded as a function of time by measuring the increase in absorbance at 350 nm. Tubulin (20 μM) was polymerized in an assembly buffer (filled circles) and in the presence of Plug-X (open circles), Plug-H (filled triangles), Plug-F (open squares) and scrambled peptides (open triangles and filled squares) at 50 μM concentrations.

their tubulin binding ability as deduced from MD simulations. The two active peptides were further tested to evaluate their effect on tubulin critical polymerization concentration. We determined the critical concentration of tubulin: a value of 9.52 μM was obtained with tubulin polymerized in the absence of peptides, but values of 14.54 μM and 13.67 μM were observed in the presence of Plug-X and Plug-H, respectively. Therefore, Plug-X and Plug-H competitively affect tubulin assembly in vitro by decreasing the elongation rate and increasing the tubulin critical concentration.

Tests were also performed on the lung adenocarcinoma A549 cell line. A proliferation assay was initially performed: values ($\pm$ s.e.m.) of 184.3 $\pm$ 12.3 μM and 197 $\pm$ 11 μM were derived for the IC50 in the presence of Plug X and Plug H, respectively. Plug-F and the two scrambled peptides showed no significant cytotoxic activity at any of the tested concentrations. In order to correlate the anti-proliferative effects shown by Plug-X and Plug-H with microtubule damage, immunofluorescence and confocal microscopy analyses on A549 cells were performed, and the microtubule network was investigated. Microtubules distribution underwent dramatic rearrangements in the presence of Plug-X, Plug-H, and thiocolchicine, a well-known inhibitor of

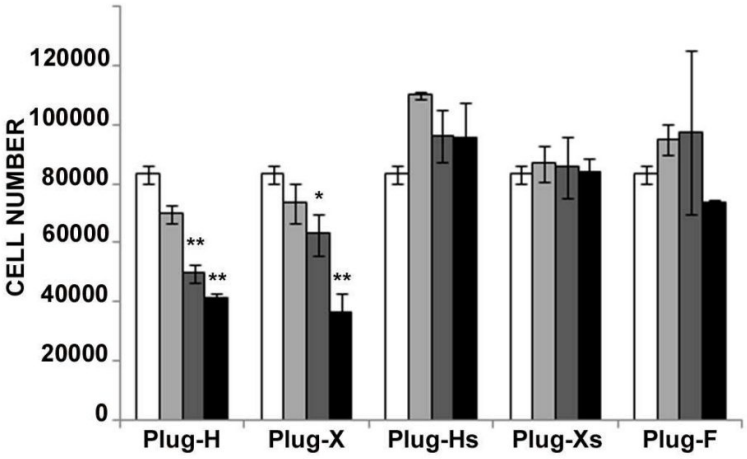

Figure 4.17: Cell proliferation assay was performed on A549 cells incubated for 48 h in the absence (white) or with increasing concentration of plugs (10 μM, light gray; 50 muM , gray; 250 muM, black).

microtubule assembly that causes the complete breakdown of microtubule network. The majority of cells treated with the peptides shrinked and showed significant microtubule damages.

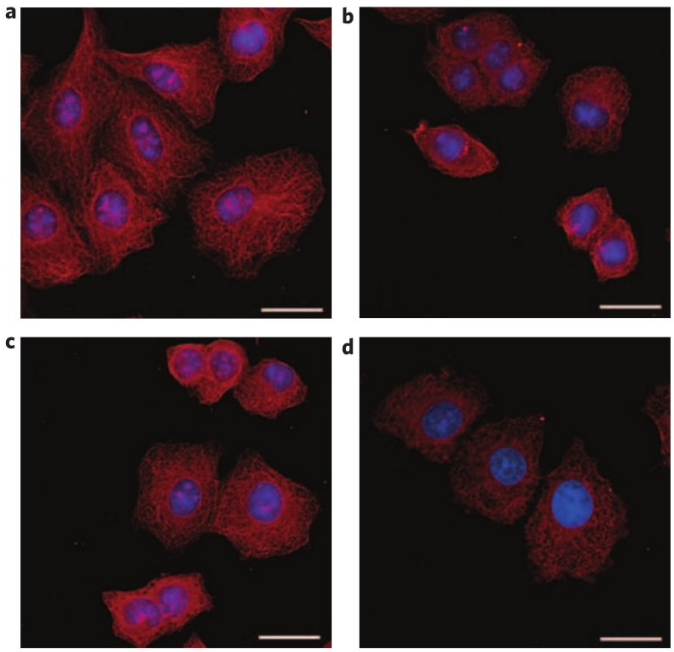

Figure 4.18: Microtubule organization in human lung carcinoma cell line A549 exposed for 24 hours to different treatments. a, Control cells. b, Cells exposed to 50 μM Plug-X. c, Cells exposed to 50 μM Plug-H. d, Cells exposed to 500 nM thiocolchicine. Microtubules are revealed by immunofluorescence localization of a tubulin (red, nuclei in blue). Scale bars are 20 μm.

Conclusions

Our approach based on an "in silico" study of PPIs resulted successful. Combining MD and CAS we were able to identify the "hot-regions" important for tubulin self-association and to design three peptides, obtained by tubulin subsequences rich in hot-spots. The MD simulations of the three peptide-tubulin complexes, predicted that two out of three peptides would retain the ability to bind tubulin even when extracted from the mother sequence. Biological tests on the peptides confirmed the computational results: Plug-H and Plug-X exhibited significant activity against tubulin polymerization and cell proliferation, thus giving an interesting contribution to the knowledge of protein-protein interactions in microtubules and, in prospect, to the development of novel antimitotic, tubulin-targeted drugs. The peptides are now object of a patent application and the succeful approach has been reported in a recent publication[62].

References for Part I

[1] Jordan, M. A. and Wilson, L. (2004) Microtubules as a target for anticancer drugs. *Nature Reviews Cancer 4*, 253–265.

[2] Amos, L. and Klug, A. (1974) Arrangement of subunits in flagellar microtubules. *Journal of Cell Science 14*, 523–549.

[3] Nogales, E., Whittaker, M., Milligan, R. A., and Downing, K. H. (1999) High-resolution model of the microtubule. *Cell 96*, 34–79.

[4] Vale, R. D., Coppin, C. M., Malik, F., Kull, F. J., and Milligan, R. A. (1994) Tubulin GTP hydrolysis influences the structure, mechanical properties, and kinesin-driven transport of microtubules. *Journal of Biological Chemistry 269*, 23769 –23775.

[5] Nogales, E. and Wang, H.-W. (2006) Structural mechanisms underlying nucleotide-dependent self-assembly of tubulin and its relatives. *Current Opinion in Structural Biology 16*, 221–229.

[6] Schek, H. T., Gardner, M. K., Cheng, J., Odde, D. J., and Hunt, A. J. (2007) Microtubule assembly dynamics at the nanoscale. *Current Biology 17*, 1445–1455.

[7] Desai, a. and Mitchison, T. J. (1997) Microtubule polymerization dynamics. *Annual Review of Cell and Developmental Biology 13*, 83–117.

[8] Mitchison, T. and Kirschner, M. (1984) Dynamic instability of microtubule growth. *Nature 312*, 237–242.

[9] Walker, R. A., O'Brien, E. T., Pryer, N. K., Soboeiro, M. F., Voter, W. A., Erickson, H. P., and Salmon, E. D. (1988) Dynamic instability of individual microtubules

analyzed by video light microscopy: rate constants and transition frequencies. *The Journal of Cell Biology 107*, 1437–1448.

[10] Mejillano, M. R., Barton, J. S., and Himes, R. H. (1990) Stabilization of microtubules by GTP analogues. *Biochemical and Biophysical Research Communications 166*, 653–660.

[11] Kirschner, M. W. (1978) Microtubule assembly and nucleations. *International Review of Cytology 54*, 1–71.

[12] Mickey, B. and Howard, J. (1995) Rigidity of microtubules is increased by stabilizing agents. *The Journal of Cell Biology 130*, 909–917.

[13] Caplow, M., Ruhlen, R. L., and Shanks, J. (1994) The free energy for hydrolysis of a microtubule-bound nucleotide triphosphate is near zero: all of the free energy for hydrolysis is stored in the microtubule lattice. *The Journal of Cell Biology 127*, 779–788.

[14] Krebs, A., Goldie, K. N., and Hoenger, A. (2005) Structural rearrangements in tubulin following microtubule formation. *EMBO reports 6*, 227–232.

[15] Wang, H.-W. and Nogales, E. (2005) Nucleotide-dependent bending flexibility of tubulin regulates microtubule assembly. *Nature 435*, 911–915.

[16] Rogers, G. C., Rogers, S. L., and Sharp, D. J. (2005) Spindle microtubules in flux. *Journal of Cell Science 118*, 1105–1116.

[17] Walczak, C. E., Cai, S., and Khodjakov, A. (2010) Mechanisms of chromosome behaviour during mitosis. *Nature Reviews. Molecular Cell Biology 11*, 91–102.

[18] Sharp, D. J., Rogers, G. C., and Scholey, J. M. (2000) Microtubule motors in mitosis. *Nature 407*, 41–7.

[19] Maiato, H., Deluca, J., Salmon, E. D., and Earnshaw, W. C. (2004) The dynamic kinetochore-microtubule interface. *Journal of Cell Science*.

[20] Biggins, S., Walczak, C. E., and Specification, K. (2003) Captivating Capture: How Microtubules Attach to Kinetochores. *Science 13*, 449–460.

[21] Giannakakou, P., Sackett, D., and Fojo, T. (2000) Tubulin/microtubules: still a promising target for new chemotherapeutic agents. *Journal of the National Cancer Institute 92*, 182–183.

[22] Checchi, P. (2003) Microtubule-interacting drugs for cancer treatment. *Trends in Pharmacological Sciences 24*, 361–365.

[23] Pasquier, E. and Kavallaris, M. (2008) Microtubules: a dynamic target in cancer therapy. *IUBMB life 60*, 165–170.

[24] Jordan, M., Thrower, D., and Wilson, L. (1991) Mechanism of inhibition of cell proliferation by Vinca alkaloids. *Cancer Research 51*, 2212–2222.

[25] Gupta, S. and Bhattacharyya, B. (2003) Antimicrotubular drugs binding to vinca domain of tubulin. *Molecular and Cellular Biochemistry* pp. 41–47.

[26] Wang, C., Ravelli, R. B. G., Roussi, F., Steinmetz, M. O., Knossow, M., and Curmi, P. A. (2005) Structural basis for the regulation of tubulin by vinblastine. *Nature 435*, 519–522.

[27] Nguyen, T. L., Mcgrath, C., Hermone, A. R., Burnett, J. C., Zaharevitz, D. W., Day, B. W., Wipf, P., Hamel, E., and Gussio, R. (2005) A Common Pharmacophore for a Diverse Set of Colchicine Site Inhibitors Using a Structure-Based Approach. *Journal of Medicinal Chemistry* pp. 6107–6116.

[28] Wang, Y., Yin, S., Blade, K., Cooper, G., Menick, D. R., and Cabral, F. (2006) Mutations at leucine 215 of beta-tubulin affect paclitaxel sensitivity by two distinct mechanisms. *Biochemistry 45*, 185–194.

[29] Ganesh, T., Guza, R. C., Bane, S., Ravindra, R., Shanker, N., Lakdawala, A. S., Snyder, J. P., and Kingston, D. G. I. (2004) The bioactive Taxol conformation on beta-tubulin: experimental evidence from highly active constrained analogs. *Proceedings of the National Academy of Sciences of the United States of America 101*, 10006–10011.

[30] Gupta, M. L., Bode, C. J., Georg, G. I., and Himes, R. H. (2003) Understanding tubulin-Taxol interactions: mutations that impart Taxol binding to yeast tubulin. *Proceedings of the National Academy of Sciences of the United States of America 100*, 6394–6397.

[31] Panda, D., DeLuca, K., Williams, D., Jordan, M., and Wilson, L. (1998) Antiproliferative mechanism of action of cryptophycin-52: kinetic stabilization of microtubule dynamics by high-affinity binding to microtubule ends. *Proceedings of the National Academy of Sciences of the United States of America 95*, 9313–9318.

[32] Dumontet, C. and Sikic, B. I. (1999) Mechanisms of action of and resistance to antitubulin agents: microtubule dynamics, drug transport, and cell death. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology 17*, 1061–1070.

[33] Hamel, E. (1996) Antimitotic natural products and their interactions with tubulin. *Medicinal Research Reviews 16*, 207–231.

[34] Risinger, A. L., Giles, F. J., and Mooberry, S. L. (2009) Microtubule dynamics as a target in oncology. *Cancer Treatment Reviews 35*, 255–261.

[35] Lo Conte, L., Chothia, C., and Janin, J. (1999) The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology 285*, 2177–2198.

[36] Chène, P. (2006) Drugs targeting protein-protein interactions. *ChemMedChem 1*, 400–11.

[37] Arkin, M. R. and Wells, J. A. (2004) Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nature Reviews. Drug discovery 3*, 301–17.

[38] Berg, T. (2003) Modulation of protein-protein interactions with small organic molecules. *Angewandte Chemie (International ed. in English) 42*, 2462–81.

[39] Janin, Y. L. (2003) Peptides with anticancer use or potential. *Amino acids 25*, 1–40.

[40] Lowman, H. B. (2003) Bacteriophage display and discovery of peptide leads for drug development. *Annual Review of Biophysics and Biomolecular Structure 26*, 401.

[41] Sidhu, S. S., Fairbrother, W. J., and Deshayes, K. (2003) Exploring protein–protein interactions with phage display. *Chembiochem 4*, 14–25.

[42] Sattler, M., Liang, H., Nettesheim, D., Meadows, R. P., Harlan, J. E., Eberstadt, M., Yoon, H. S., Shuker, S. B., Chang, B. S., Minn, A. J., and Others. (1997) Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. *Science 275*, 983.

[43] Kussie, P. H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A. J., and Pavletich, N. P. (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science 274*, 948.

[44] Zhong, H. and Carlson, H. A. (2005) Computational studies and peptidomimetic design for the human p53–MDM2 complex. *Proteins: Structure, Function, and Bioinformatics 58*, 222–234.

[45] Villacanas, O. and Rubio-Martinez, J. (2006) Reducing CDK4/6-p16INK4a interface. Computational alanine scanning of a peptide bound to CDK6 protein. *Proteins: Structure, Function, and Bioinformatics 63*, 797–810.

[46] Adcock, S. A. and McCammon, J. A. (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical Reviews 106*, 1589–1615.

[47] Frenkel, D. and Smit, B. (2002) *Understanding molecular simulation: from algorithms to applications*. (Academic Press), p. 672.

[48] Allen, M. P. and Tildesley, D. J. (1993) *Computer simulation of liquids*. (Oxford University Press), p. 412.

[49] Gilson, M., Given, J., Bush, B., and Mccammon, J. (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophysical Journal 72*, 1047–1069.

[50] Swanson, J. M. J., Henchman, R. H., and McCammon, J. A. (2004) Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophysical journal 86*, 67–74.

[51] Beveridge, D. L. and DiCapua, F. M. (1989) Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annual Review of Biophysics and Biophysical Chemistry 18*, 431–492.

[52] Straatsma, T. P. and McCammon, J. A. (1991) Multiconfiguration thermodynamic integration. *The Journal of Chemical Physics 95*, 1175.

[53] Massova, I. and Kollman, P. A. (2000) Combined molecular mechanical and continuum solvent approach {(MM-PBSA/GBSA)} to predict ligand binding. pp. 113–135.

[54] Bashford, D. and Case, D. A. (2000) Generalized born models of macromolecular solvation effects. *Annual Review of Physical Chemistry 51*, 129–52.

[55] Cunningham, B. and Wells, J. (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science 244*, 1081–1085.

[56] Moreira, I. S., Fernandes, P. A., and Ramos, M. J. (2006) Computational Alanine Scanning Mutagenesis - An Improved Methodological Approach. *Journal of Computational Chemistry 28*, 644–654.

[57] Massova, I. and Kollman, P. A. (1999) Computational Alanine Scanning To Probe Protein-Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *Journal of the American Chemical Society*.

[58] Huo, S., Massova, I., and Kollman, P. A. (2002) Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *Journal of Computational Chemistry 23*, 15–27.

[59] DeLano, W. L. (2002) Unraveling hot spots in binding interfaces: progress and challenges. *Current opinion in structural biology 12*, 14–20.

[60] Moreira, I. S., Fernandes, P. A., and Ramos, M. J. (2006) Unraveling the importance of protein-protein interaction: application of a computational alanine-scanning mutagenesis to the study of the IgG1 streptococcal protein G (C2 fragment) complex. *The Journal of Physical Chemistry. B 110*, 10962–9.

[61] Löwe, J., Li, H., Downing, K. H., and Nogales, E. (2001) Refined structure of alpha beta-tubulin at 3.5 A resolution. *Journal of Molecular Biology 313*, 1045–1057.

[62] Pieraccini, S., Saladino, G., Cappelletti, G., Cartelli, D., Francescato, P., Speranza, G., Manitto, P., and Sironi, M. (2009) In silico design of tubulin-targeted antimitotic peptides. *Nature Chemistry 1*, 642–648.

Part II

Osmolytes and Protein Folding

INTRODUCTION

Harsh Environments and Protein Stability

The chemistry that rules proteins interactions is a fascinating subject, unveiling slowly in recent years. How macromolecules are able to react to different environment and to interact with chemical compounds, exogenous or endogenous, still preserving a well defined biological activity, selectivity towards their targets and stability, remains a masterpiece of balance between different chemical, and physical, forces, acting at the same time[1, 2]. A balance that, sometimes, is achieved in a bewildering simple way. Indeed, to protect proteins, and cells in general, from severe environmental conditions, such as extreme temperatures, pressures or salinity levels, living organisms had devised a rather plain response strategy: increasing the level of small organic compounds, called osmolytes[3–5]. Osmolytes had been observed in a wide range of organisms, from bacteria to mammals, with variable concentrations and effects[6–8] and accumulation of them had been found in some species, called "extremophiles", able to survive in extremely harsh conditions, as in deserts[9], deep waters[10, 11] or icy environments[12]. Several barophilic bacteria, as Psychromonas hadalis, for example, proliferate optimally at pressures of 50 MPa[13–15] while Methanopyrus kandleri, a hyperthermophilic organism, can grow at temperatures as high as 120°C[16, 17]. Unveiling the molecular mechanisms responsible for this uncommon adaptability is of primary importance especially under an agricultural perspective: drought tolerance, for example, as showed by some angiosperms[18, 19], the notorious resurrection plants,

able to survive under severe drought, and to re-flourish when water is available again, could be a relevant factor in overcoming the difficulties that still afflict crop production.

Figure 4.19: The so-called "resurrection plants" are able to recover from complete dessication and flourish when water is available.

Osmolytes, and their effects, are well known and documented since many years, starting from the seminal studies by Yancey et al., analysing the distribution of these molecules in different organisms and connecting their uptake to the prevention of water loss as a consequence of high saline concentrations. Since those first evidences, the protective role of osmolytes had been widely confirmed and demonstrated. For example, it had been demonstrated that, while as an early response to the increase of NaCl external concentration, bacteria quickly accumulate K^+ ions[20, 21], a slower concomitant accumulation of organic osmolytes, especially Glycine Betaine (GB), is observed[22]. The more "compatible" small organic solutes progressively substitute the positive ions in the cellular environment, thus constituting a slower, but more tolerable and persistent, counteraction to osmotic stress. Moreover, it had been proved that GB synthesis, generally occurring by means of glycine metilation by S-adenosylmethionine, is a K^+ dependent reaction, that lead to osmolyte accumulation only for K^+ concentrations greater that 0.4 M, hence correlating, on a metabolic level, GB production to positive ions uptake[23]. Proteins stability at high temperatures had also been connected to accumulation of osmolytes, as trehalose, able to protect

from thermal denaturation, extending the range of temperatures in which the native state is energetically favored[24]. In relatively recent years, Bolen and co-workers had carefully characterized thermotolerance enhancement effects from a thermodynamic point of view[25, 26], thus confirming the alteration of proteins free energy landscapes and of folded-unfolded states equilibrium. They also succeeded in demonstrating that the stabilization goes beyond plain prevention of denaturation: in their study[27], Trimethylamine N-oxide (TMAO) could properly fold even highly unfolded proteins, resulting from destabilizing mutations, to conformations that acquire biological activity and a proper native-like structure. Beside protecting osmolytes, also denaturing osmolytes exist, promoting protein unfolding, as the well known urea.
Interestingly, despite the extremely wide ensemble of proteins in living organisms, only a few dozen osmolyte molecules exist, results of an accurate genetic convergence, suggesting that their action must be as universal as possible. To sustain these arguments, it had also been demonstrated that osmolytes are interchangeable and that artificially discontinuing the availability of one of them, could lead to accumulation of other alternative solutes[28], as for sorbitol in rat renal medulla, whose inhibition triggers a compensating increase in glycine betaine. Besides, protective effects had been proved to be exerted even by exogenous osmolytes, neither produced nor used naturally by an organism, once they are added in the cellular environment[29].

Notwithstanding the properties and peculiarities highlighted so far, these constitute evidences of the protecting (or denaturing) behavior of osmolytes, but give no clue about how these molecules are able to carry out their function. The mechanism behind osmolytes-mediated phenomena is still matter of debate. Much remains to be learned about these metabolites and a satisfactory and complete description of the phenomenon is still to be achieved.

Since atomic-level simulations can provide accurate and detailed informations on these molecules interactions, we decided to study the effects of protecting and denaturing osmolytes on model proteins in order to shed more light on the osmolytes mechanism of action.

*In the following, our approach to the study of osmolytes will be presented. Firstly, a brief description of the most common osmolytes and of their applications will be given (**chapter 5**). An overview of the different theories proposed so far for the mechanism of action will also be reported in the last section. In **chapter 6**, the problems connected with osmolytes simulations, and in particular with the reproduction of protein folding, will be discussed and a summary of the available computational techniques to tackle this problem will be presented. Finally (**chapter 7** and **8**) results of our simulations will be presented for two different proteins, a small $\beta-$hairpin fragment and a prevalently helical mini-protein.*

CHAPTER 5

Osmolytes

Osmolytes are simple organic molecules, generally non-toxic so that they do not perturb cell machinery even in high concentration, to the extent that they are commonly referred to as "compatible solutes". They can be divided into a few major categories[3, 30]: small carbohydrates including sugars (e.g. trehalose), polyols (glycerol, inositols, sorbitol, etc.) and derivatives (such as o-methyl-inositol); amino acids (glycine, proline, taurine, etc.) and derivatives (e.g. ectoine); methylamines (such as trimethylamine N-oxide (TMAO) and glycine betaine) and methylsulfonium solutes including dimethylsulfonopropionate (DMSP); and urea.

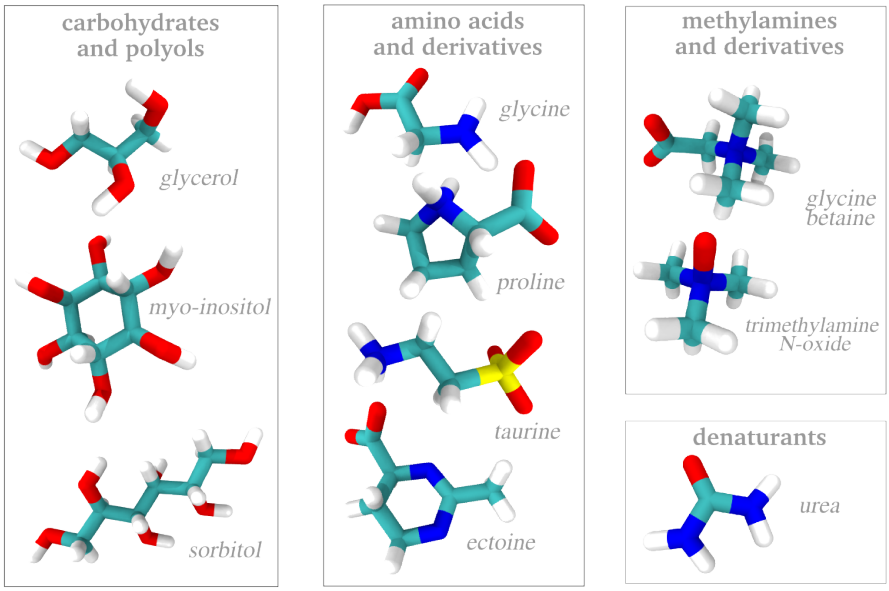

Figure 5.1: Some of the most common osmolytes, such as Glycine Betaine (GB), Trimethylamine N-oxide (TMAO), Taurine and the denaturant Urea.

Except for urea (used only by relatively few types of animals), these categories are

widespread in occurrence. For example, **plants** experiencing salt stress accumulate osmolytes as proline, ectoine, betaine, glucose, or myo-inositol in the cytoplasm of their cells[31]. Their accumulation is commonly connected to an increase of the rate at which they are synthesized, but the mechanisms by which the increase is triggered is known only in a few specific cases. Examples include mannitol, synthesized via the action of a mannose-6-phosphate reductase, in celery, or tomato plants that, as a response to an increase of salinity, enhance sucrose concentration by an enhancement of sucrose-phosphate synthase activity. In plants, as in bacteria, glycine betaine is synthesized by oxidation of choline[32] and its accumulation in leaves of stressed plants is regulated, at least partially, through changes in the expression of genes for the biosynthetic enzymes. However, the signals that provoke these changes in gene expression have not been identified.

Marine species living at shallow depths, generally accumulates betaine, several taurine derivatives, trimethylamine N-oxide, glycine[33] or scyllo-inositol[34].

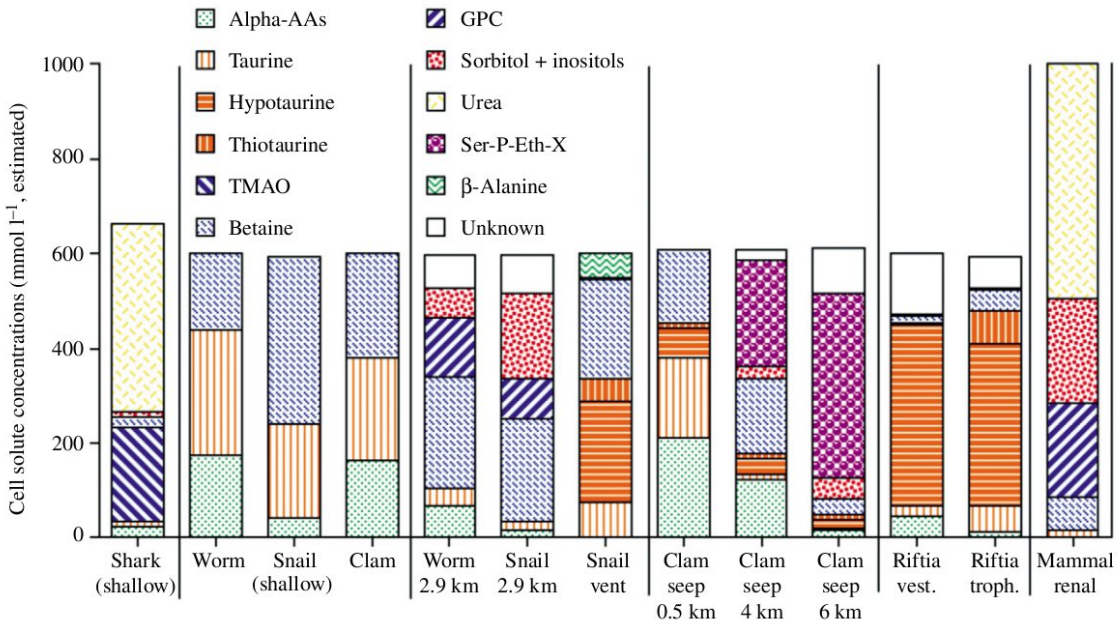

Figure 5.2: Distribution of osmolytes in different species, from bacteria to mammals.

Interestingly, trimethylamine N-oxide increases with depth of habitat in muscles of some species, apparently counteracting the perturbing effects of hydrostatic pressure[34]. Some other species accumulate urea along with methylamines in their

tissues and extracellular fluid to balance the high osmolality of sea water[3]. Even if urea denatures proteins and alters their enzymatic activity, methylamines that accumulate with urea seem to counteract its denaturing effects.

In ***mammals***, the osmolality of blood is normally kept remarkably constant and contained; therefore, most mammalian cells are not normally exposed to the extreme osmolalities experienced by the cells discussed above[35]. Nevertheless, certain mammalian cells contain considerable concentrations of organic osmolytes, and most of the rest are able to accumulate them if suitably stressed. Renal medullary cells contain the highest levels of osmolytes as a consequence of the extremely high concentrations of NaCl and urea connected with their function[36]. The principal organic osmolytes in renal medullary cells are sorbitol, betaine, inositol, taurine, and glycerophosphocholine (GPC).

5.1 Practical applications of osmolytes

The properties of osmolytes are becoming[37, 30] increasingly useful in molecular biology, agriculture and biotechnology. It was suggested that stabilizing osmolytes, sometimes called "chemical chaperones", might be able to rescue misfolded proteins involved in human diseases[38]. Indeed, recently, it was proved that addition of various mammalian osmolytes and TMAO can indeed restore function of one form of cystic fibrosis mutant protein[39] and that TMAO can prevent misfolding of prion proteins[40]. Some of these solutes, especially taurine and sometimes inositol and glycine betaine, are major ingredients of a number of energy or sport drinks. However, many of these solutes have unique metabolic reactions and could cause harmful side-effects if used where their non-osmotic properties are not needed. As anticipated, osmolytes have interesting potential application under an agricultural perspective. These molecules are potentially able, indeed, to mitigate the effects of drought and soil salinization, the most prominent factors in limiting agricultural productivity worldwide. More than 20% of all cultivated lands are estimated to be salt-stressed with up to one-third of agricultural lands being salt-affected in certain countries[37]. Drought and limited water availability, even for a short period, can cause several damages and have a deleterious impact on cultivated lads and, at a higher level, on economy. As a

consequence of global warming and climate instability, drought events are becoming more frequent (see Figure 5.3).

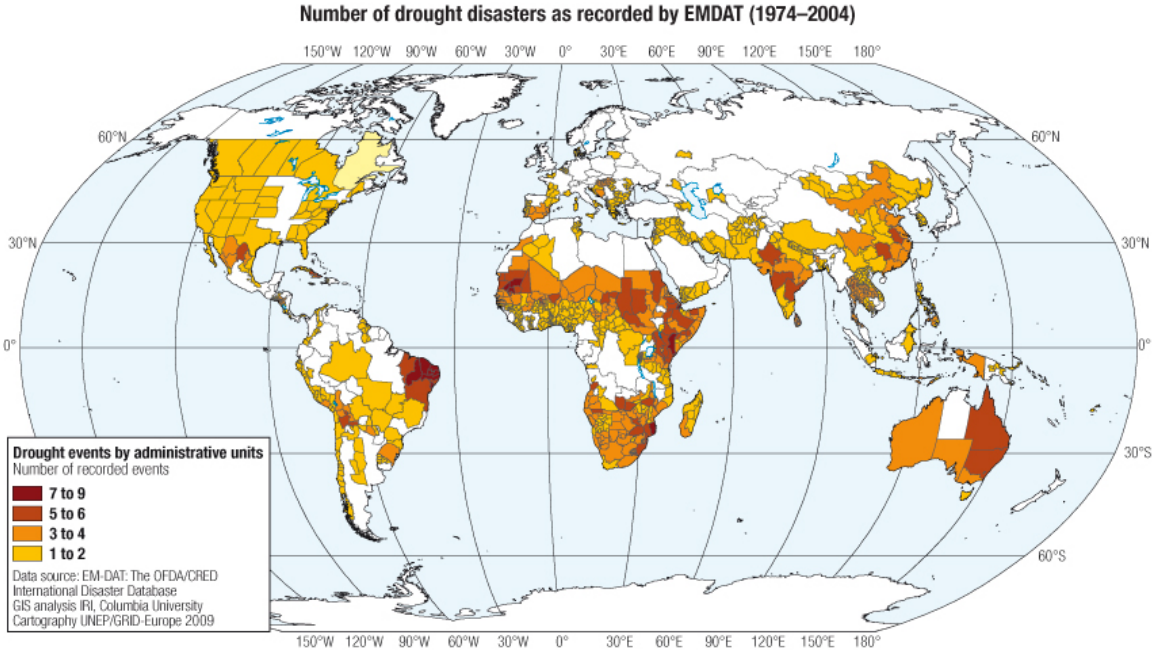

Figure 5.3: The number of "drought disasters" occurred in the last decades. The damages caused by water deficiency can have a negative recoil on economy.

From 1950, most land areas have warmed up by 1-3°C, with the largest warming over northern Asia and northern North America[41].

Unfortunately, attempts to generate plant varieties with improved salinity or drought tolerance using selection-based breeding strategies have proved largely unsuccessful[42, 43] due to the well-recognized complexity of salinity and drought tolerance traits. Attempts to engineer improved tolerance using gene transfer are limited by an incomplete understanding of stress tolerance mechanisms. For these reasons, and for to their diverse function roles, osmolytes have become the favorite tools for the development of improved osmotic stress tolerance. In particular, identification and utilization of novel osmoprotectants derived from stress-tolerant organisms will aid in such improvements[44]. Combinatorial engineering of more than one osmoprotectant should also improve the efficacy of these approaches by reducing the metabolic costs of osmolyte production. Nevertheless, defining the exact mechanisms of protection and the specific macromolecules being protected will undoubtedly lead to further improvements in **osmolyte-mediated protection strategies**.

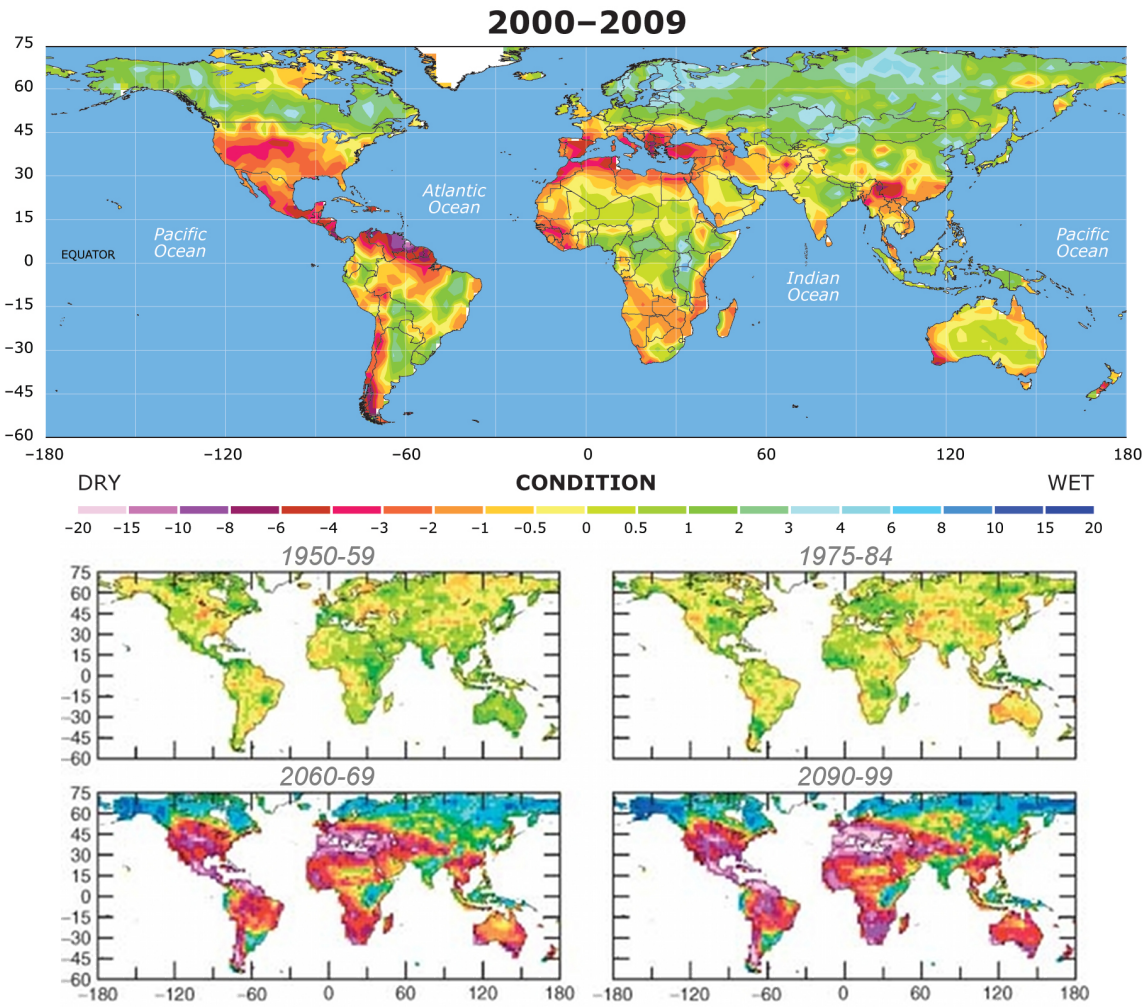

Figure 5.4: Current trend in water availability according to Palmer Drought Severity Index (PDSI). Red to pink areas are extremely dry (severe drought) conditions while blue colors indicate wet areas relative to the 1950–1979 mean. Adapted from Dai et al.[41].

5.2 Direct or Indirect Mechanism?

As discussed in previous section, many benefits will derive from a deep understanding of osmolytes protecting effect. Unfortunately, little is known about osmolytes mechanism of actions. So far, two different mechanisms, **direct** and **indirect**, have been proposed both for protecting osmolytes and for urea; while direct and indirect effects have been observed and discussed extensively in the literature, especially for urea, no decisive and indisputable conclusion has been made[45–50].

A large number of experiments have been devoted to gain a deeper knowledge of osmolytes mechanism. Infrared experiments have indicated that urea has little

effects on the water structure, while protecting osmolytes as TMAO, on the other hand, significantly change the water spectra[51]. NMR experiments[52, 53] on protein structures have revealed that the addition of urea to the protein solution resulted in a more flexible protein backbone, in particular the N-H bond, while osmoprotectants reduced this flexibility and made the protein backbone more rigid. These studies suggested that direct interactions, possibly through hydrogen bonds between urea and the protein backbone, are likely responsible for the denaturing effects of urea, especially considering its negligible effects on water structure and dynamics[48, 54–56], while stronger indirect effects may be responsible for the protecting effect of osmoprotectants. The denaturation of a number of proteins and polypeptides by urea studied using molecular dynamics simulations[57–60] has not been conclusive on whether direct or indirect effect is the main driving force for protein denaturation. In agreement with an indirect mechanism, urea was observed to "free" water molecules from its hydrogen-bonding network, which then interacts more strongly with the protein backbone[48], while direct interaction of urea with both the backbone and the polar residues and/or the apolar side chains were also observed[56, 59]. Simulations by Zou et al.[47] are consistent with ***transfer free energy*** studies, in which the free energy to transfer a protein (or a residue) from water to an osmolyte solution is evaluated. According to transfer free energies results, protecting osmolyte are excluded from the protein surface (an effect termed "osmophobic effect") while urea strongly interacts with the protein backbone. The main effect on the relative stability between the native and unfolded states arises then from favorable/unfavorable interactions with the protein backbone, with osmoprotectants disfavoring the unfolded state. The studies also showed that osmolytes significantly changs water structure and hydrogen bonding, consistent with an indirect effect. However, other simulation studies questioned the existence of the indirect effect on water structure[61, 62]. Another important question in understanding the effects of cosolvents on protein structures has been on the side-chain interactions. Probably partly due to the cancellation of hydrophobic and hydrophilic effects[61] and due to the weakness of each individual factors, the effects of osmolytes on the hydrophobic interactions have been found small by simulations[61, 62]. The contribution of side chains in transfer free energy was also found to be small[63–65] compared to the backbone transfer energy.

Because of the complexity in the interactions between water, cosolvent, and solute, the molecular characterization of the "indirect" and "direct" effects on protein structure remains unclear and questionable. Apparently, a unified atomic-level picture of osmolytes behavior and on their direct/indirect effects on protein stability is still missing.

CHAPTER 6: COMPUTATIONAL METHODS II

The Limits of Molecular Dynamics

In order to study osmolytes mechanism of action, we need to be able to reproduce accurately the effects of these molecules on protein stability. This means that we need to be able to reproduce folding and unfolding phenomena for the proteins studied, in order to evaluate how these processes are affected by the protecting or denaturing osmolytes. Unfortunately, while *in silico* experiments have a potential to investigate in details processes at an atomic level, their use can be computationally expensive, due to the complexity of the algorithms and functions needed to reproduce a system realistically. As already mentioned, integration steps, used to evolve the system, has to be of the order of magnitude of the fastest dynamics, i.e. approximately of the order of 1 fs. As a result, the time scales that can be afforded with the available computational power are restrained to a limited range, at best, approaching a few microseconds. The large conformational rearrangements, as those involved in protein folding and protein-protein association, can easily take place on time scales that are orders of magnitude larger. Hence, nowadays, they still constitutes *rare events* for molecular dynamics simulations.

Alternatively, we might settle for an approximate calculation of the free energy of unfolding, in order to check that this ΔG_{unfold} is higher in the osmolytes solution, i.e. the denaturation of the protein is disfavored due to an increased stability. This would be the same approach used for protein-protein interactions, where the problem of the long time scales needed for protein-protein association was bypassed calculating "a posteriori" a binding free energy. Naturally, this is a challenging task due to the

problems connected with free energy calculations that have been extensively discussed in **Computational Methods I**. Moreover, the approximated solutions provided by end-points methods as MM-GBSA are unfeasible when dealing with unfolding, since this unimolecolar processes involve large and progressive conformational transitions. Actually, the problems connected with time-scales and the ones linked to free energies are the two sides of the same coin, since a correct reproduction of the process and an effective sampling of the conformations involved will also provide the informations needed to calculate free energy. Indeed, if all the relevant conformations r_i contributing to $U(r_i)$ are explored with a frequency that is proportional to their probability, free energy can easily be computed "counting" how many times a region of the conformational space has been explored, following so-called "weighted histogram" approaches.

Drawing on these considerations, it is clear that in the case of folding not only MD cannot be used to reproduce the process, but it is not suitable to provide an approximated ΔG either: the use of more advanced techniques becomes mandatory. Numerous accelerated-MD variants have been developed in order to extend the conformational sampling and the accessible time scales, allowing to simulate "rare events". A brief overview will be presented in the remainder of this chapter.

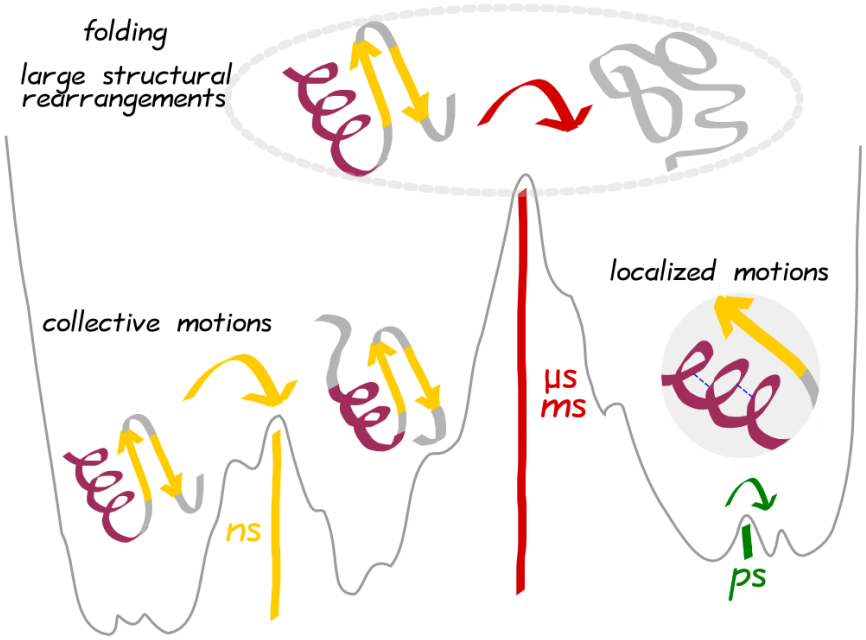

Figure 6.1: The free energy landscape of a protein is generally rugged and presents various minima. While several motions can be sampled using MD, the sampling of large structural rearrangements, as those involved in conformational transition, is usually beyond the reach of molecular dynamics.

Accelerated MD methods can be loosely grouped into three broad classes: (1) the first class alters sampling of conformational space though **explicit modification of the potential surface**; (2) the second class also alters the sampling but by using non-Boltzmann sampling to **increase the probability of high-energy states** and (3) the third class includes those methods that **enhance the sampling of certain degrees of freedom** at the expense of other, typically faster, degrees of freedom. Naturally, there is a certain degree of overlap between these classes and other classification are equally popular. For example, distinctions can be made regarding the scope and applicability of these methods, ranging from those aimed at reconstructing a low-dimensional projection of the free energy landscape to those focused on the exploration of reactive transitions and saddle points. In the whole, there are two primary goals driving the development of enhanced sampling methods. Some of these approaches aim to increase the volume of conformational space that is explored during the simulation, while others aim to drive the system to a particular conformation or to the global minimum energy conformation more rapidly.

6.1 Overcoming the Limits of MD

Perhaps the most obvious approach for enhancing sampling of high- energy states is to raise the simulation temperature. This approach is known as high-temperature dynamics and has been evaluated by several groups[66–68]. It is suggested that high-temperature molecular dynamics is a useful aid in conformational searches, but physiologically relevant low- energy structures are not generally obtained even after minimization of the generated high-energy structures. Indeed, the generated structures often have an infeasible proportion of cis-peptides[66]. Consequently, this method is not without criticism. Another possibly major issue is that MM force fields, generally, have not been designed for, or validated with, temperatures much beyond the physiologically relevant 300-330 K that most MD simulations are run at. Whether or not the force field is physically correct, there may be debate about whether the use of such high temperatures leads to appropriate sampling characteristics since the entropic contribution to the free energy is significantly enhanced (and therefore over-sampled)[69]. Although there are several problematic issues involved with using high temperatures to accelerate MD simulations, the basic principle acts as the fundamental basis for a few more complex approaches discussed below.

6.1.1 Modified Potentials

The basic principle behind various potential-energy modification methods is to reduce the amount of time that the simulated system remains in a local energy minimum well, speeding the transitions from the region of one local minimum to another, forcing the system to sample the remainder of the available conformational space. In each of these approaches, the potential-energy function is altered to enhance sampling by reducing the propensity of energy wells to act as conformational traps. Methods that modify the potential-energy surface include the deflation method[70], conformational flooding[69, 71], umbrella sampling[72], local elevation[73], potential smoothing[74], puddle-skimming and and puddle-jumping method[75], hyperdynamics[76], accelerated MD[77] or diffusion equation method (DEM)[78]. The smoothing or flattening of the potential-energy landscape can be applied either globally to increase the overall sampling or only along a specific, predefined, reaction coordinate to enhance the conformational

evolution along the direction of a desirable transformation or enhance sampling over a particular set of conformations, depending upon the method. For example, given some prior information about the desired conformations, one widely used approach is umbrella sampling[72]. A compensating function, known as an umbrella potential, is added to the potential-energy function to bias the sampling. Obviously, construction of the umbrella function requires prior knowledge of the conformations of interest. In principle, these schemes are all applicable to methods other than MD and can be used for enhancing sampling by MC simulations, for example. The recently described accelerated molecular dynamics[77] allows for more rapid sampling of the configurational space in systems with rough energy landscapes and also allows one to calculate the correct thermodynamic properties of the system. In the accelerated MD method, time becomes a statistical quantity in the simulation. The effective time scale of simulations is increased by several orders of magnitude at the expense of sampling around the energy minima. The local elevation method[73] enhances sampling by adding a penalty potential to any conformations previously sampled. This approach has not been adapted for use with proteins. Similarily, in conformational flooding[69, 71] the initial state of the system is destabilized by adding an extra unfavorable potential at this initial state. A Gaussian potential is added to the system to force it along PCA-selected essential degrees of freedom. Unlike local elevation, this method has been demonstrated in protein simulations. On a similar idea is based the recently developed **Metadynamics** technique, that will be extesively described in the following.

6.1.2 Modified Sampling

The basic principle behind most of the second class of methods is also to reduce the amount of time that the simulated system remains in local energy minimum wells, forcing the system to sample the remainder of the conformational space available. However, these techniques use alternative methods for sampling rather than performing conventional MD on explicitly modified energy landscapes. Modified sampling methods include high-temperature MD discussed above, locally enhanced sampling (LES)[79],replica exchange[80, 81], parallel tempering[82, 83], self-guided MD[84], milestoning[85], and various non-Boltzmann sampling methods[86, 87]. Several approaches that use a series of simultaneous (or parallel) MD simulations are

demonstrated in the literature[79, 88, 89]. These tend to be very well suited to trivial parallelization with communication only between the individual systems. Indeed, the well-known distributed protein-folding project, Folding@Home, uses a multiple-copy approach[90]. Like the multiple-copy approaches, **Replica Exchange Molecular Dynamics** (REMD)[81] and the closely related parallel tempering method[83] utilize a series of simultaneous and noninteracting simulations, known as replicas. With proteins, these simulations are typically MD, but earlier work applied Monte Carlo simulations. The replicas are simulated over a range of temperatures, and at particular intervals the temperatures of these simulations may be swapped (i.e., replicas are exchanged). The methods differ, however, in the way the individual simulations are coupled. Usually these replicas may exchange temperatures according to Monte Carlo-like transition probability. Such exchanges occur through a simple swapping of the simulation temperatures via velocity re-assignment. The high-temperature replicas jump from basin to basin, but the low-temperature replicas explore a single valley with sampling characteristics just like conventional MD. While REMD is widely applied to smaller molecules, particularly in peptide and protein folding experiments[91, 92], it is found to be extremely computationally expensive when applied to large proteins.

In a study in which REMD was applied to a 20-residue peptide it was found that at physiologically relevant temperatures the conformational space was sampled much more efficiently than it was with conventional constant temperature MD[92] and with similar thermodynamic properties.

6.1.3 Modified Dynamics

This third class of enhanced sampling method encompasses those methods in which the dynamics along the "slow" degrees of freedom are accentuated relative to the "fast" degrees of freedom. One such method is extremely widely used, to the extent that simulations not applying it (or one of its close descendants) are exceptionally rare. This is the SHAKE algorithm[93], where constraints are applied to particular bond lengths to allow larger time steps to be taken without encountering excessive forces. This basic idea is pushed further in **Coarse-Grain** (CG) simulations where a system is represented with a reduced number of degrees of freedom compared to

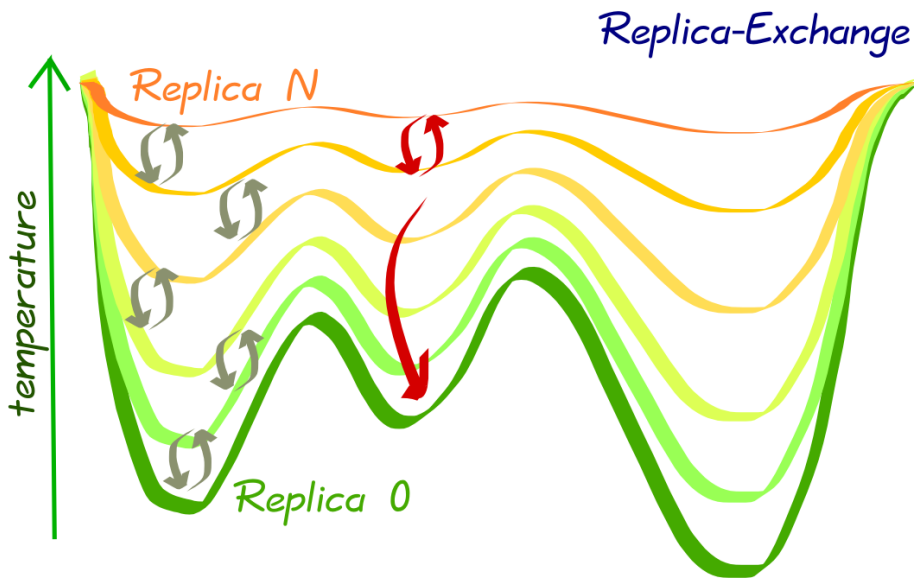

Figure 6.2: Pictorial representation of Replica Exchange Molecular Dynamics (REMD). Since the different replica (O to N) have different (increasing) temperatures, they will experience a progressively smoother potential. Through a chain of exchanges, conformations that are only accessible to the high temperature replicas can be explored also by low temperature (i.e. target) replicas.

its all-atom description. Due to this reduction, and consequently to the elimination of fine interaction details, the simulation of a CG system requires fewer resources and an increase of several order of magnitude in the simulation time and length scale can be achieved. However, the most important contribution to the speed up of the CG simulations is not generally due to the reduced number of particle, and hence of interactions, but to the removal of the fastest dynamics, allowing the use of a much larger integration step. Generally, CG models are the grouped into three main classes according to the number of interacting centers for each amino acid. Passing from the coarsest (one bead) to the finest (four-six beads) class, the explicit representation of the sidechain and peptide bond atoms is progressively included[94].

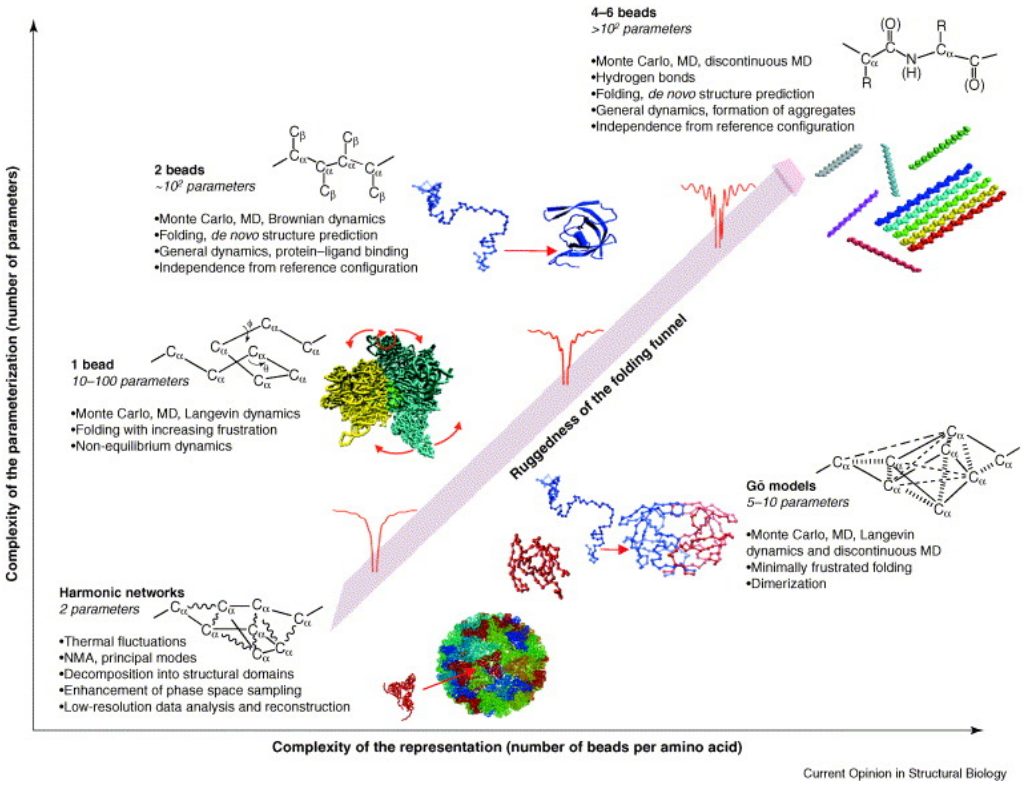

Figure 6.3: Representation of the features of Coarse Grain models, the complexity increases with the number of beads

6.2 Metadynamics

The Metadynamics technique, developed by Parrinello & Laio in 2002, combines many features of other techniques described above. First of all, the algorithm is based on "dimensional reduction", i.e. it requires the identification of a proper set of Collective Variables (CV) which are assumed to be able to describe the process of interest. The dynamics in the CV-space, defined by the chosen CVs, is biased by the addition of a potential whose main peculiarity is that of being "history-dependent". Indeed, the potential is not defined preliminary as in Umbrella-sampling and derivative methods, but is constructed progressively as a sum of small gaussians. The energy hills are added in the already sampled points of the CV-space, an thus progressively disfavour already sampled conformations. The sum of the gaussian, interestingly, can also be exploited to obtain an estimator of the free energy. In detail, the system is described by a set of n CVs $S_n(x)$, explicit functions of the coordinates x. The probability distribution as a function of the values $s = (s_1...s_n)$ of a CV can hence be written as

$$P(s) = \frac{exp(-\frac{1}{T}F(s))}{\int ds\ exp(-\frac{1}{T}F(s))}$$

and the free energy $F(s)$ is given by

$$F(s) = -T\ ln\left(\int dx\ exp\left(-\frac{1}{T}V(x)\right)\delta(s - S(x))\right)$$

If a trajectory $x(t)$ can be obtained for a very long time, then the probability at time t can be calculated as

$$P(s) \sim \frac{1}{t}\int_0^t dt'\ \delta(S(x(t')) - s)$$

i.e. taking the hystogram of the CV along the trajectory. In a conventional MD simulation s will be confined to a local minimum of $F(s)$, with a low probability of escaping from it. If we add an external potential of the form $\frac{k}{2}(s - S(x))^2$, then the system will be pushed to explore the region around s, even if this is not a minimum of $F(s)$. For example, we can force the system to explore all the possible values of s in a systematic way, keep changing the value of s in our potential. Ideally, the process would be more efficient if this value of s can change "naturally" throughout the simulation, as a consequence of the system dynamics. Metadynamics acts exactly like that. Every τ_G steps of MD, a energy hill is deposited in the explored region, constituting a little repulsive potential in the present value of s. Progressively, the repulsive hills will sum up and disfavour the present region of the CV-space, pushing the system towards unexplored regions. The external potential, "history-dependent", at a time t will then be

$$V_G(S(x), t) = w \sum_{\substack{t'=\tau_G, 2\tau_G \ldots \\ t'<t}} exp\left(-\frac{(S(x) - s(t'))^2}{2\delta s^2}\right)$$

where $s(t) = S(x(t))$ is the value of the CVs at time t and w, δs and τ_G are respectively the height, the width and the deposition frequency of the gaussians. At a generic

time t the system will be in a point $s(t) = S(x)$ of the CV-space and subjected to all the gaussian positioned in different points $s(t')$ at a previous ($< t$) time t' that, since gaussian are added every τ_G, must be a multiple of τ_G. Naturally, since w, δs and τ_G determine the amount of energy deposited they are deeply connected with both the speed of the free energy "filling" and the error in the determination of $F(s)$. In detail, while the speed is proportional to the three parameters, it has been derived that the error $\varepsilon(s)$ is defined by the equation

$$\bar{\varepsilon}^2_{approx} = C\frac{S^2 wT}{D\tau_G}\frac{\delta s}{S}$$

where C is a constant, typical of the dimensionality of the CV-space and D is the diffusion coefficient.

6.2.1 Metadynamics-derived Methods

Metadynamics had been widely recognized as an efficient tool for the study of biomolecules, however, since the definition of a small number of CVs is needed, the risk of neglecting relevant (slow) degrees of freedom is a major concern. Moreover, using a large number of CVs too avoid under-sampling, or because the relevant variables are not just two or three, can lead to a significant worsening of the performance, since the filling of a many-dimension free energy surface would be computationally expensive, and would need a significant amount of time for the bias to converge. Several extensions of the metadynamics methodology have been proposed, most of which based on the use of several replicas, in order to alleviate this drawback.

Parallel Tempering Metadynamics

Parallel Tempering Metadynamics (PTmetaD) is based on the idea of running several metadynamics in parallel. It is based on the combination of metadynamics with parallel tempering, the most popular replica-exchange implementation (see section 6.1.2 on pag. 83). The simulation of the different replica at different temperatures allows to collect informations from different metadynamics and to improve the statics, leading to an indirect sampling of the neglected CVs. In the spirit of RE, an exchange

of coordinates of two replicas at adiacent temperatures is attempted every τ_x steps. The acceptance ratio takes into account the fact that different replicas experience different bias potentials. The acceptance ratio for an exchange involving replicas i and j will be

$$P = min\Bigg\{1, exp\Bigg[\left(\frac{1}{T_j} - \frac{1}{T_i}\right)(U(r_j) - U(r_i)) + \frac{1}{T_i}(V_i(s(r_i)) - V_i(s(r_j))) + \\ + \frac{1}{T_j}(V_j(s(r_j)) - V_j(s(r_i)))\Bigg]\Bigg\}$$

where U is the ordinary potential, r_i and T_i are the coordinates and the inverse temperatures of replica i and V_i is the bias potential of replica i before the exchange. If the move is accepted, the coordinates are exchanged and the momenta are rescaled as

$$r_i' = r_j \qquad p_i' = \sqrt{\frac{T_i}{T_j}}p_j$$

$$r_j' = r_i \qquad p_j' = \sqrt{\frac{T_j}{T_i}}p_i$$

Since the free energy profiles are filled in parallel at all temperatures, the dynamics of the system rapidly becomes diffusive in the CV-space.

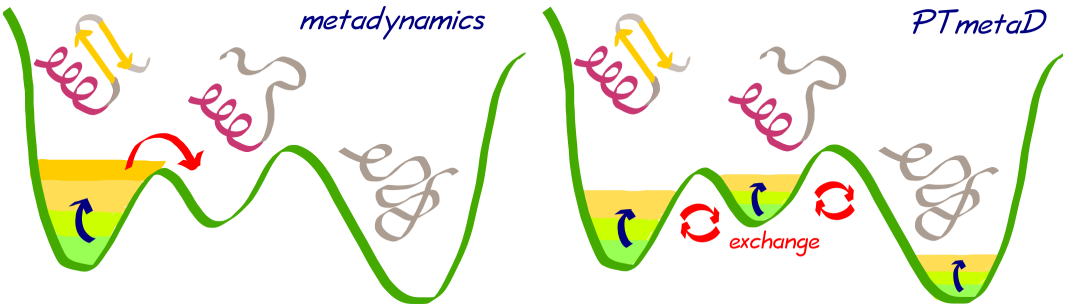

Figure 6.4: Comparison of Metadynamics and Parallel Tempering Metadynamics (PTmetaD). Instead of filling progressively the different minima of the surface (a), the exchange of coordinates allows a parallel filling of all the basins.

While metadynamics take advantage of the combination with PT because it improves the sampling of degrees not explicitly included in the CVs, also PT is enhanced since

metadynamics provide a better and more reliable sampling of the barriers. Indeed, the barriers, being less populated, are generally under-sampled in PTm while the flattening of the free energy surface due to the metadynamics potential hills favours an exploration of all the relevant barrier conformations.
The main drawback of PTmetaD is inherited from parallel tempering and is the relevant computational cost, due to the different replica running in parallel. To cope with that, metadynamics can also be combined with variants of PT that are less computationally expensive, such as Solute Tempering (ST).

Bias Exchange Molecular Dynamics

The Bias-Exchange MD approach is based on the same idea of the aforementioned PTmetaD, i.e. using RE to enhance sampling, but connects the two techniques in a different way. In BEMD, instead of perfoming different metadynamics at different temperatures, one runs a large number of metadynamics at the same temperature, biasing each replica with a potential acting only on just one or two CVs. If all the CV used are relevant, the various replica biased along one or two of them would be affected by relevant errors. However, since the different replica can exchange coordinates in BEMD, each trajectory evolves in the high-dimensional CV-space, "sequencially" biased by one or two CVs, allowing to explore efficiently the free energy landscape. This time, the exchange move between replica i and j, with coordinates x_i and x_j, involves the swapping of the potentials $V_G(S_k(x_i), t)$ and $V_G(S_l(x_j), t)$

$$P = min\left\{1, exp\left[\frac{1}{T}V_G(S_k(x_i), t) + V_G(S_k l(x_j), t) - V_G(S_k(x_j), t) - V_G(S_l(x_i), t)\right]\right\}$$

Opposed to PTmetaD, BEMD does not eliminate the necessity to define proper CVs, allowing "indirect sampling" of neglected degrees of freedom, but allows to use a much higher number of CVs, without deleterious effects on the performance of metadynamics, since only low-dimensional projection are filled with gaussians and not the overall high-dimensional surface in which the system evolves thank to the exchange of bias. However, since the number of CV can be larger, pretty simple and

standard CV can be used, each describing only small changes in the conformations, because every single CV doesn't need anymore to properly describe the entire process of interest alone.

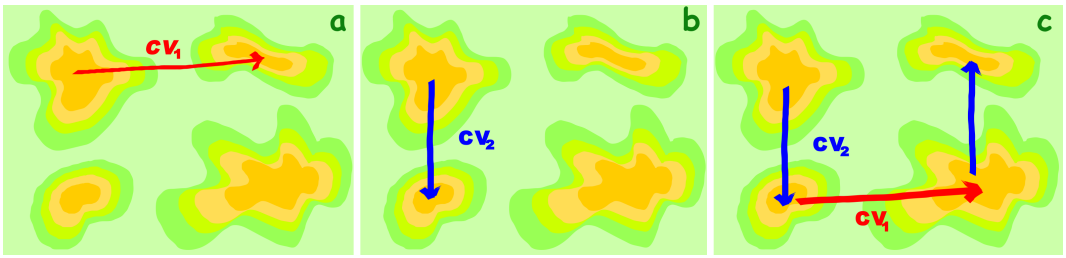

Figure 6.5: Pictorial explanation of BEMD. If the bias is applied along CV1 (a) or CV2 (b) alone only some minima can be reached. If exchange of bias between replica can take place, conformations for which both CV1 and CV2 change can be explored. While for a small number of CV this can also be done applying the bias along both CV1 and CV2, for a large number N of CV this will affect performance because a N-dimensional surface need to be filled. The exchange of bias allows to do that while filling mono(bi)-dimensional landscapes.

CHAPTER 7

First studies on a β-hairpin fragment

Among the different techniques reviewed in the previous chapter, we selected **metadynamics** as the most suitable one, being relatively cheap from a computational point of view. Moreover, this techniques allows to obtain free energy landscapes in a straightforward manner, thanks to the history-dependent bias.

The first study was carried out using a small model system, the Protein G, B1 domain $\beta-$hairpin. The GB1 $\beta-$hairpin has a fairly simple structure, two strands connected with a turn, and is among the most widely studied peptides[95, 96], with experimentally determined kinetic and thermodynamic properties. Several computational studies had addressed the problem of $\beta-$hairpin folding mechanism by means of long MD simulations[97], Monte Carlo techniques[98] and replica exchange molecular dynamics (REMD)[99, 100]. Metadynamics studies on this system had also been carried out extensively[101, 102]. Moreover, $\beta-$hairpin has an intrinsic interest due to its relative stability at room temperature[95] even as a separate fragment and because it is one of the first elements that is formed during secondary structure organization, thus being an ideal model for the study of folding. Effects of the osmolyte on the $\beta-$hairpin free energy surface are investigated and melting temperature variations in solutions of GB and urea are evaluated. Since metadynamics adds a non-Markovian energy term, further investigation at the molecular level were achieved performing conventional MDs on the conformations observed in the free energy landscape.

7.1 Metadynamics simulations

Metadynamics simulations were performed on three systems: β-hairpin in water, in 1 M solution of GB and in 1 M solution of urea. Since the attention is focused on GB protecting effects at a temperature close to the calculated melting one[99], metadynamics were carried out at 350 K.

COMPUTATIONAL DETAILS

The model β-hairpin fragment of G protein was obtained from the entire G protein PDB structure (PDB code: 2GB1)[103]. Three different systems had been set up: protein in water (control), protein in 1M GB and protein in 1M urea. Metadynamics simulations were performed using GROMETA 2.0[104], a modified version of GROMACS 3.3.3[105], with OPLSAA force field[106]. The systems were solvated with TIP3P water molecules[107] in a cubic box with a side of 50.0 Å. A sufficient number of waters was replaced with GB or urea molecules to reach the correct concentration, in all simulations except the control one. Three minimizations were performed, with a steepest descent algorithm, minimizing, respectively, the protein alone, the mixed solvent, and finally the entire system. Two different 50000 step equilibrations were performed, the first being NVT, the second NPT. NPT conditions were then applied for all the production phase, using a temperature of 350 K and Nose'-Hoover thermostat[108]. Integration step was 2 fs. Particle Mesh Ewald method[109] was used for long range electrostatic interactions along with a 0.8 nm cutoff.

The simulations allowed to obtain the Free Energy Surface (FES) as a function of the chosen Collective Variables (CVs) (Figure 7.1) from the history-dependent bias introduced by metadynamics.

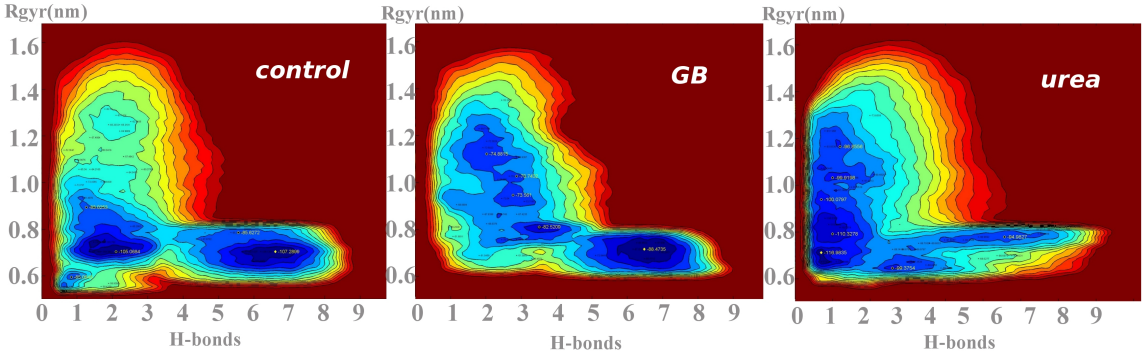

Figure 7.1: Metadynamics FES obtained at 350 K for the three systems as a function of the chosen CVs, hydrogen bonds number (horizontal axis) and gyration radius (vertical axis). Contours are spaced every 7 kJ/mol. (a) The control simulation showed two well-defined minima in the folded and unfolded basins, reporting comparable energies. (b) Metadynamics in 1 M GB showed a considerable stabilization of the folded basin compared to the homogeneously populated unfolded region . (c) Metadynamics in 1 M urea showed a considerable decrease of the folded basin population and several highly stable unfolded minima.

COMPUTATIONAL DETAILS

The CVs were chosen accordingly to a previous work by Parrinello et al.[101] as the most suitable for the discrimination of folded and unfolded structures. Hence, the number of hydrogen bonds in the peptide backbone is calculated as

$$N_H = \sum_{i \in H} \sum_{j \in O} \frac{1 - (r_{ij}/d_0)^6}{1 - (r_{ij}/d_0)^{12}}$$

with $d_0 = 2.3$ Å, taking into account only the relevant ones, i.e. those with an odd, larger than four, sequence separation[101]. The radius of gyration is computed as

$$R_{gyr} = \sqrt{\sum_i \left(r_i - \frac{1}{N_a} \sum_j r_j \right)^2}$$

with summations running over all the N_a heavy atoms. The height of the gaussians is 1.0 kJ/mol, while the width is 0.1 along hydrogen bonds number and 0.01 along gyration radius. Gaussians are deposited every 100 steps. Each metadynamics run was scheduled for 25 ns, but considered concluded upon convergence of the free energy difference between folded and unfolded state for more than 5 ns.

In order to evaluate the effects of the two co-solvents on the β-hairpin dynamics, the ΔG of denaturation was calculated as

$$\Delta G_{unfold} = k_B T \, log \left(\frac{\int_F ds \, exp\left(-\frac{G(s)}{k_B T}\right)}{\int_U ds \, exp\left(-\frac{G(s)}{k_B T}\right)} \right) \quad [1]$$

were F and U denote, respectively, the regions of the CV space corresponding to the folded and the unfolded basins. A value of 2.19 kJ/mol was calculated for the control simulation, revealing that the temperature chosen was indeed quite close to the melting temperature as expected. The positive sign of the free energy suggested that the process is disfavored in water, meaning the folded structure is the most stable. The ΔG for the simulation with GB was 9.78 kJ/mol, a considerable increase in respect to the control simulation, meaning that the osmolyte stabilized the folded conformations. On the other hand, the metadynamics in urea reported a ΔG of -18.5 kJ/mol, reporting a change of sign in respect to the control simulation, thus correctly reproducing the denaturing effect of urea.

7.2 Further Simulations

Considering that the choice of the CVs highly affects the results of a metadynamics run, and that eventual neglect of relevant degrees of freedom can lead to bias results, hybrid simulations that combine metadynamics with replica exchange algorithms had been advised[101]. Among them, Solute Tempering (ST) algorithm was chosen, since it allows to use a considerably lower number of replicas and a higher ΔT.

7.2.1 ST-MetaD simulations

The FES obtained at 350 K, with the corresponding minima structures, are reported in Figures 7.2—7.4, while FES for the first four temperatures are reported in Figure 7.5.

COMPUTATIONAL DETAILS

Solute Tempering Metadynamics were performed on the same systems using 10 temperatures in geometrical progression in the range 300 K - 507 K. This time, the height w of the gaussians for the replica m was rescaled with the corresponding temperature T_m, according to $w_m = 1.0 k_B T_m$. Swapping between neighboring replica was attempted every 0.2 ps. The average acceptance rate for the Monte carlo-like configuration swap, verified a posteriori, fell in the ideal range between 30% and 40%. All other parameters were maintained unaltered with respect to the first metadynamics runs.

The ST/metadynamics FES confirmed the results of previous metadynamics simulations: considering the replica with the same temperature used for metadynamics (i.e. 350 K) the ΔG for denaturation resulted increased upon addition of GB and decreased upon addition of urea. The ST/metadynamics in 1 M GB reported a ΔG of 13.78 kJ/mol, revealing only a slight increase in comparison with the value reported by metadynamics. Also the simulation in 1 M urea revealed little variation, showing a ΔG of -16.42 kJ/mol. Finally, the ΔG for the control simulation resulted consistent with the previously obtained one, with a slightly lower value of -5.24 kJ/mol.
By and large, the results can be considered in good agreement. The morphology of the FES, moreover, showed considerable analogies with the surfaces obtained with metadynamics. The simulation in pure water showed two almost isoergonic basins for the native and denatured structures, with the typical partially folded and misfolded conformations prevailing on completely open structures in the unfolded ensemble, in agreement with what observed in a recent paper by *Parrinello et al.*[102]. Indeed, denatured structures had either a very low gyration radius (as the typical molten globule structure) or a very low number of hydrogen bonds; completely destructured, high gyration radius, conformations corresponded to basins with significantly higher energy. An esteem of the activation energy, assuming a two state mechanism, was calculated to be 11.3 kJ/mol. In the 1M GB solution, the native minimum is considerably more stable and embraces a larger part of the CV space, meaning that slightly distorted hairpins are stabilized together with the native one; no denatured state had a energy comparable to the native β-hairpin and even

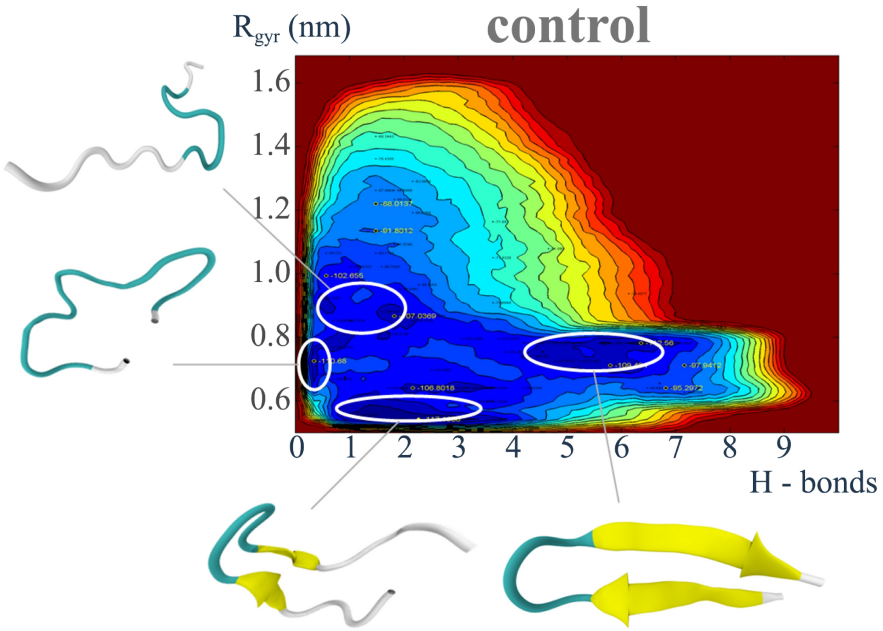

Figure 7.2: Free energy surface for ST/metadynamics at 350 K in pure water. The most relevant minima are highlighted with the respective structures.

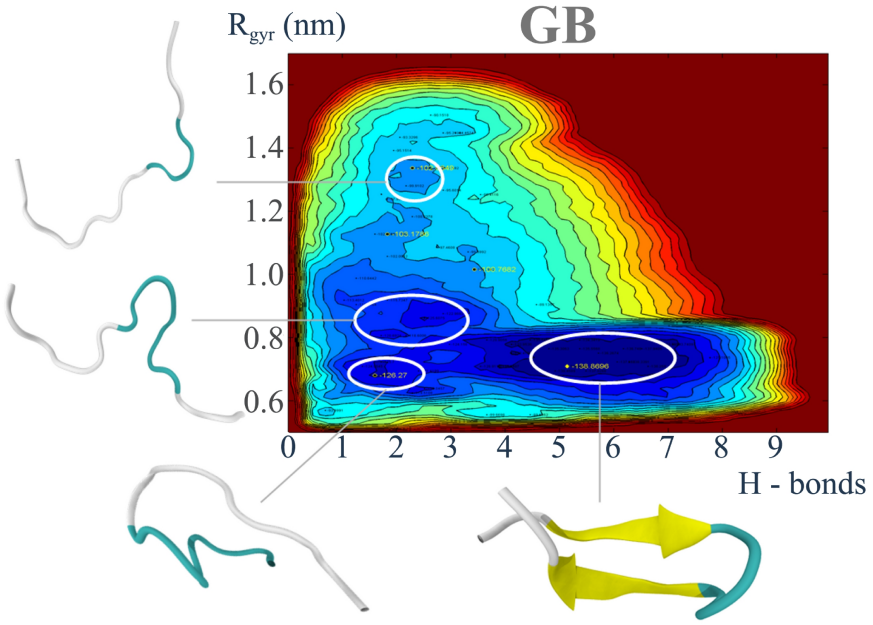

Figure 7.3: Free energy surface for ST/metadynamics at 350 K in 1 M GB. The most relevant minima are highlighted with the respective structures.

the most stable unfolded conformations reported a higher number of hydrogen bonds in respect to pure water denatured structures, retaining a gyration radius similar to

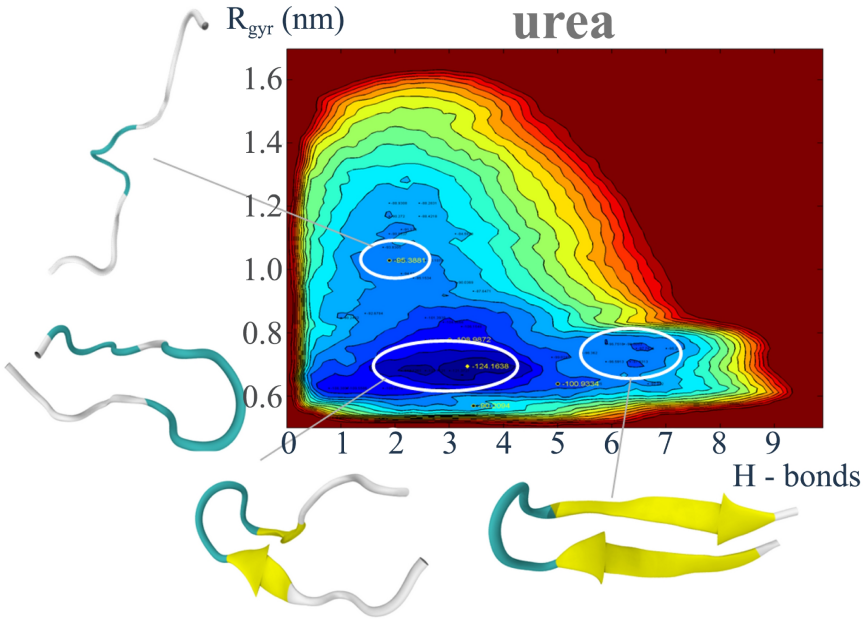

Figure 7.4: Free energy surface for ST/metadynamics at 350 K in 1 M urea. The most relevant minima are highlighted with the respective structures.

the native state, and indicating a higher level of residual structure. Noticeably, a concomitant increase of the activation energy is observed, with an estimated value around 19.9 kJ/mol. The 1M urea solution, on the other hand, presented a considerable increase of the native state energy and a prominent unfolded minimum that, once more, lies in a low gyration radius area of the CV space. Interestingly, there was substantially no energy barrier (less than 4.5 kJ/mol) for the evolution of native structure into the denatured state. ST/metadynamics also allowed to calculate the melting temperature T_m of the β-hairpin, in the approximation of a two state folding mechanism, considering the temperature dependence of ΔG. The calculated T_m for the control simulation is found to be 340 K, apparently much higher then the reported experimental values of 297 K[95]. However, previous studies[99, 110] had reported that a systematical overestimation of the melting temperature, highly dependent on the force-field parameters, is common to all computational results, giving values ranging from 330 K to 485 K. Considering the ensemble of computational results, our estimation could be considered consistent with previously reported data. In addition, the obtained T_m is consistent with previous results[91] obtained with the same force field used here. Calculation of the melting temperature in 1 M GB solution led to a

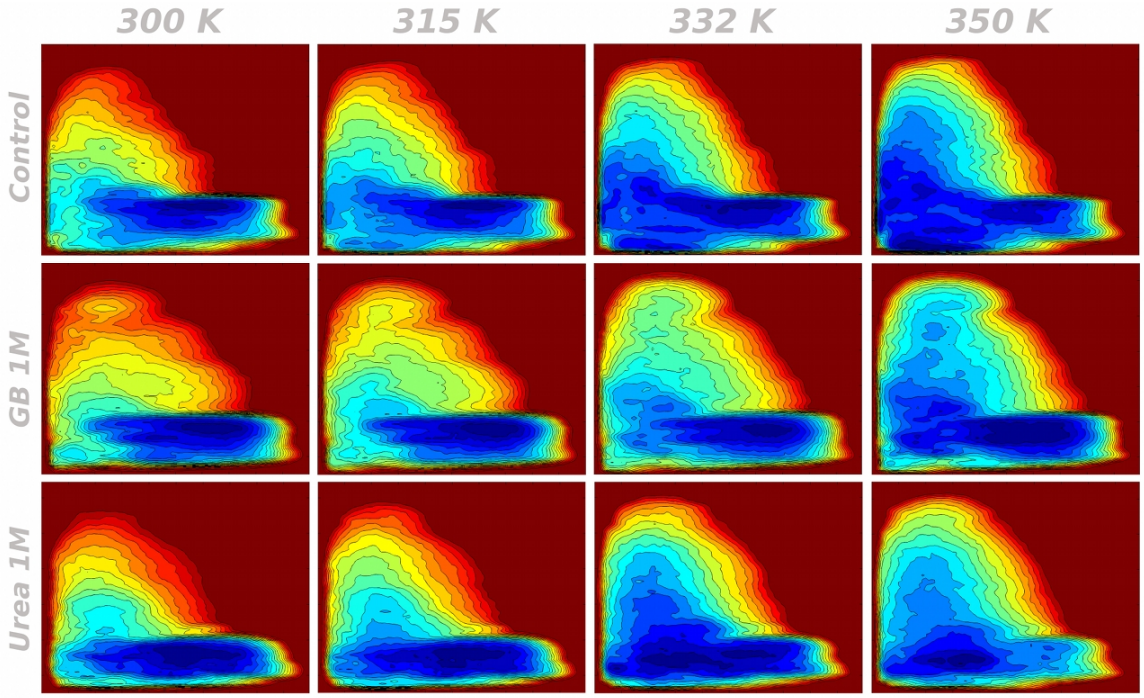

Figure 7.5: Free energy profiles for the combined ST/metadynamics simulations as a function of hydrogen bonds number (horizontal axis) and gyration radius (vertical axis). Contours are spaced every 7 k_BT, values are in kJ/mol. The four columns corresponds to the first three replica of the systems, i.e. to the first four temperatures. The three rows corresponds to the three systems simulated: control, 1 M GB and 1 M urea.

value of 384 K, more than forty degrees higher than the one reported for the control. ST/metadynamics in urea, instead, reported a T_m of 329 K, confirming the opposite behavior of the two co-solvents. The moderate decrease, in respect to GB induced increase, is consistent with urea being used as a chemical denaturing agent in much higher concentrations, up to 10 M.

7.3 Molecular properties of the mixed solvent

Considering the non-equilibrium nature of metadynamics simulations, due to the addition of the history-dependent bias, analyses at an atomic level of co-solvents configurations could not be performed accurately. A set of standard MDs was then used to investigate how the thermodynamic differences are transposed at an atomic level . Folded structures representative of the native basin were extracted from the trajectory of the metadynamics runs for every simulation.

COMPUTATIONAL DETAILS

MD simulations of the most favorable folded/unfolded conformations were performed with the same parameters of the metadynamics runs. In order to better sample the bulk properties, at a relevant distance from the protein, for preferential coefficient evaluation, the protein has been re-solvated in a larger box of 75 Å. All the MD runs were carried out for 11 ns.

Preferential coefficients and solvent density functions (SDF) were calculated to investigate the co-solvents arrangement around the protein. Since a well defined solvent bulk domain is needed for these evaluations, the selected system conformations were re-solvated to obtain a larger 75 Å box, in order to describe more properly the solvent properties at a relevant distance from the protein and to guarantee that bulk properties are constant and do not depend on the distance from the protein. During re-solvation, additional water molecules were replaced with GB or urea to obtain a slightly higher concentration (2 M). Since the ratio between water and osmolytes molecules is crucial for the calculations, higher concentration simulations resulted being more suitable, for statistical reasons, reporting more stable bulk averages. The calculated SDF for the mixed solvent simulations are reported in Figure 7.6.
The co-solvent distribution around the protein is slightly above bulk average for all GB simulations, showing a broad slowly decaying peak; no accumulation of GB molecules in the proximity of the protein is thus showed. On the contrary, urea simulations showed a prominent peak at 1.8 Åwith a protein/bulk distribution ratio of more than 3.2, meaning the density of urea is larger in the protein domain. The preferential coefficient Γ_{XP}, indeed, confirmed the results of SDF analyses: the calculated values for GB simulations are negative (-1.28) confirming the osmolyte exclusion from the protein surface, while urea simulations reported high positive values (6.84) .

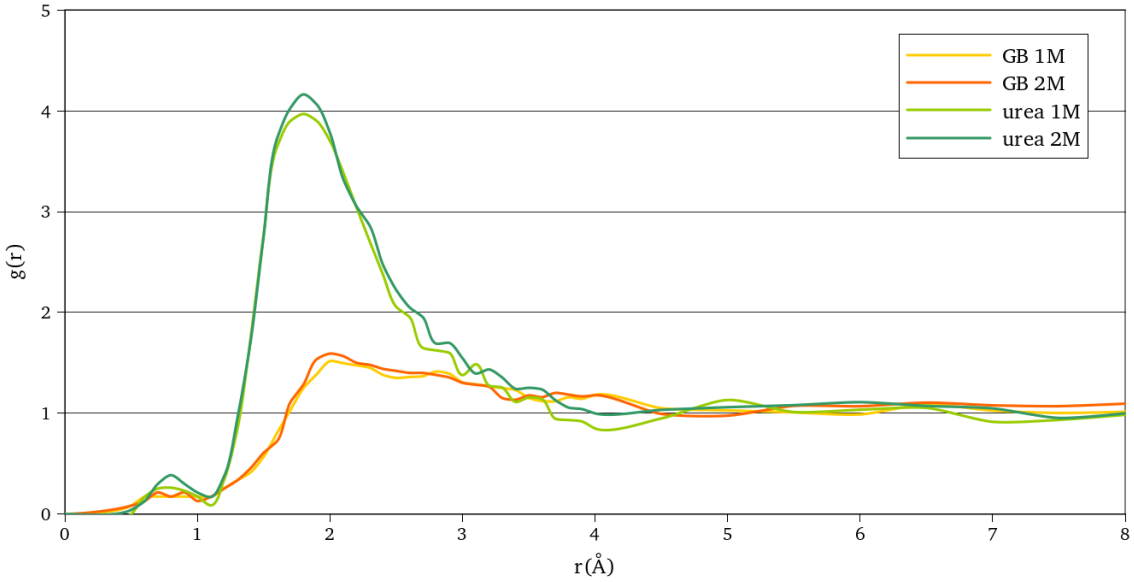

Figure 7.6: Solvent density function averaged over 500 snapshots of MDs simulations. In solution of GB, no relevant peak is showed, revealing an exclusion from the protein surface. Urea, on the other hand, showed a prominent peak at 1.8 Å.

COMPUTATIONAL DETAILS

The solvent density function (SDF)[111] describes how the molecule of osmolyte are distributed around the protein. It is, in principle, equivalent to the radial distribution function, but takes into account the shape and volume of the protein. The SDF is computed as

$$\rho_X(r) = \frac{X(r, r')}{V(r, r')}$$

where r is the radius of the solvation shell, $X(r, r')$ is the number of X molecules found from r to r' and $V(r, r')$ is the volume of the shell from r to r'. The number of molecules was obtained calculating for different bulk/protein separations, i.e. different r values. The volume $V(r, r')$ was calculated on the basis of the grid-based solvent-accessible methodology by Daggett et al.[112]. The cutoff used for molecules subdivision in the two domains during preferential coefficients calculations was chosen considering the distance required for the SDF to approach a unitary value (i.e. bulk distribution), according to previous works[112].

COMPUTATIONAL DETAILS

Preferential coefficients (Γ_{XP}) were calculated using the approach proposed by Baynes & Trout [111, 113, 114] on the basis of previous works[115, 116]. According to this theory, Γ_{XP} can be evaluated defining two domains, a bulk domain (I) and a protein one (II), and calculating

$$\Gamma_{XP} = \left\langle n_X^{II} - n_W^{II} \left(\frac{n_X^I}{n_W^I} \right) \right\rangle$$

where $n_{W/X}$ are the number of water (W) or osmolyte (X) molecules in the I/II domain. The molecules were subdivided into the different domains comparing their distance from the protein van der Walls surface with the chosen cutoff (4.5 Å).

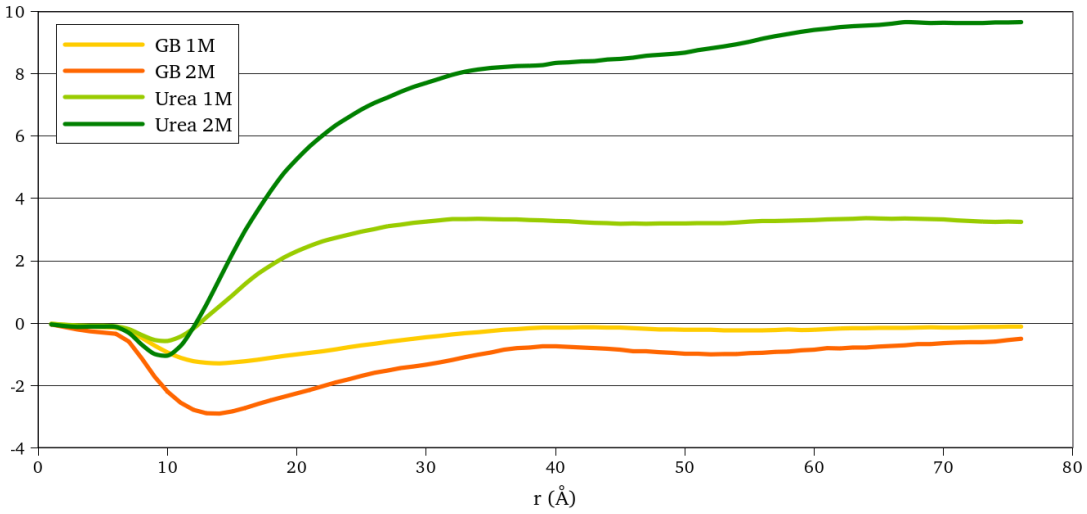

Figure 7.7: Preferential coefficient for the GB and urea solutions as a function of the cut-off for the two domains (bulk and protein).

In-depth analyses of the co-solvent distributions revealed a further difference between the two considered solutions (Figure 7.8).

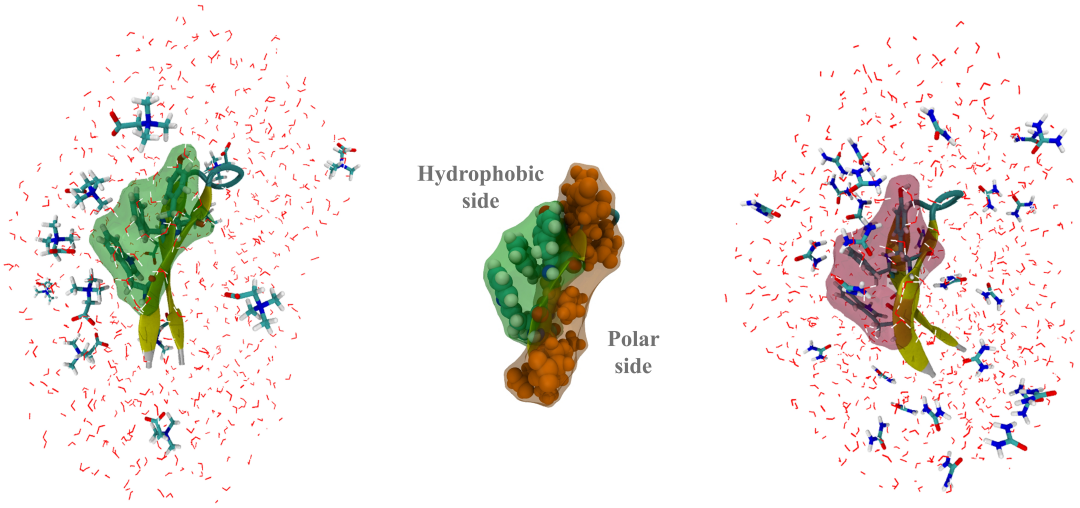

Figure 7.8: Preferential distribution of GB (left) and urea (right) around the hydrophobic core. The different polarity of the two sides of the hairpin is highlighted in the central panel. GB showed significant anisotropy in the distribution, being excluded by the majority of the residues, but interacting strongly with hydrophobic groups forming the core.

While urea distributes in a homogeneous way around the protein, GB has a marked anisotropic distribution and seemed to make contacts with the hydrophobic core, consisting of residues Trp_{43}, Tyr_{45} and Phe_{52}. The hairpin, indeed, seemed to have two opposite sides (see Figure 7.8, central panel): an hydrophobic side, where the side-chains of residues Trp_{43}, Tyr_{45} and Phe_{52} clusterize and a more hydrophilic side, with charged residues side-chains. GB prevalently accumulates on the first side, orientating the three methyl group towards the cluster and keeping the carboxyl group turned towards the solvent. The preponderance of GB contacts with hydrophobic residues Trp_{43}, Tyr_{45} and Phe_{52} is highly interesting since these exact residues had been reported to have a relevant role in β-hairpin stabilization[117]. The formation of the hydrophobic core involving these residues (Figure 7.9) had been described as the driving force for folded structure formation and is the fulcrum of many theories about hairpins folding mechanisms, mainly based on so-called hydrophobic-collapse[118, 98, 100, 119].

Moreover, these residues are the ones reporting the most negative values of experimental transfer free energy according to *Bolen et al.*[120], i.e. the transfer of these residues from pure water to the osmolyte solution is highly favorable. These

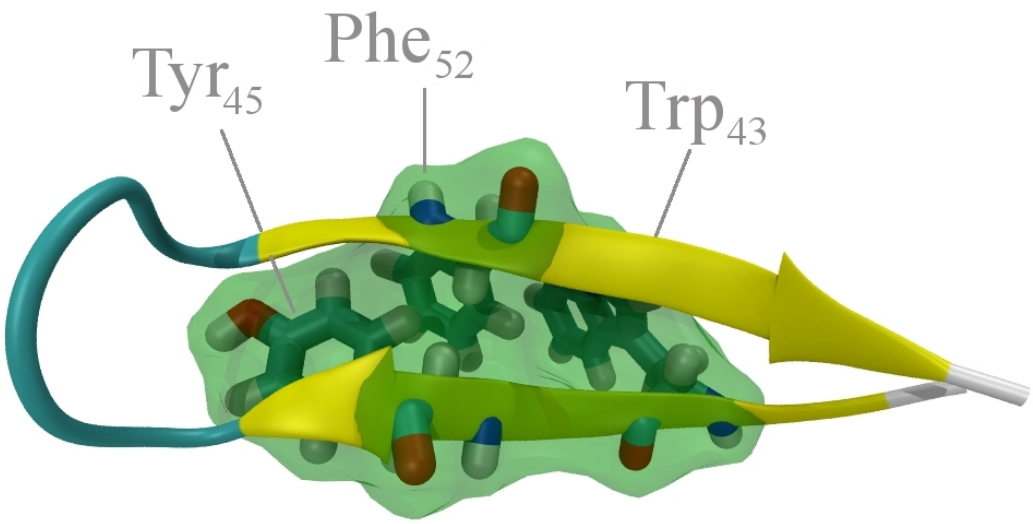

Figure 7.9: The hydrophobic core formed by Trp_{43}, Tyr_{45} and Phe_{52}. GB seemed to interact strongly with the residues of this core, giving the highest Γ_{XP} values.

features are confirmed by calculation of the preferential coefficient per residue, Γ_{XP} (Figure 7.10).

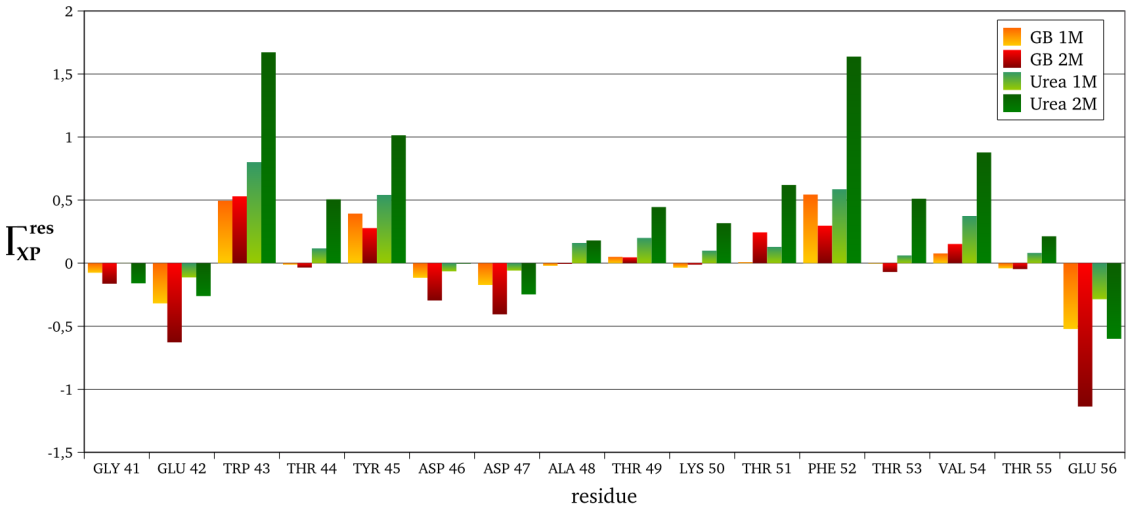

Figure 7.10: Preferential coefficient per residue. Values are generally positive for urea and negative, or close to zero, for GB. In the whole, hydrophobic residues are the one with the highest Γ_{XP}, while charged residues always preferred interactions with water molecules. It can be inferred that the osmoprotectant is not excluded from hydrophobic residues (as Trp_{43} and Phe_{52}) as opposed to the rest of the peptide.

While, on the whole, Γ_{XP} results confirmed the exclusion of GB from the protein surface, large positive per residue values were obtained for Trp_{43}, Tyr_{45} and Phe_{52} in all GB simulations, confirming that the osmoprotectant is not excluded from these residues as opposed to the rest of the peptide. This was interesting since a prevalent apolar character of the solvent accessible surface area (SASA) is a constant for most of the osmolytes, as extensively discussed in a recent paper by *Street et al.*[121], and is always in conjunction with a less prominent, generally negative, polar SASA. Urea, on the other hand, despite reporting the highest Γ_{XP} for hydrophobic residues as GB simulations (Figure 7.11), clearly showed higher values also for all the other residues, highlighting its preferential interaction with the protein.

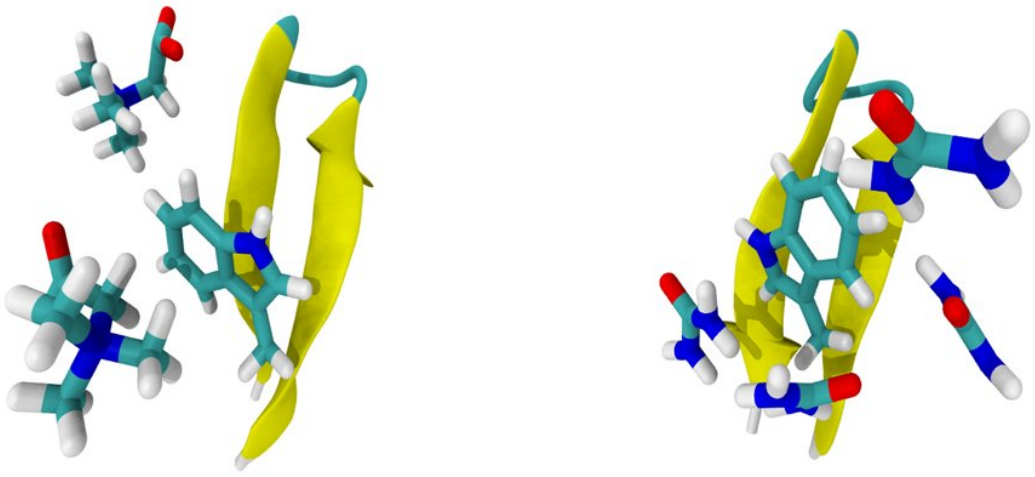

Figure 7.11: The highest number of protein/co-solvent contacts are registered for Trp43. Interaction of GB (left) and urea (right) with this residue are highlighted in a representative snapshot.

For both urea and GB, lowest values are reported for charged residues, meaning that, as expected, they always preferred interactions with water molecules. Interestingly, results for urea were in excellent agreement with previous ones by *Stumpe et al.*[55]: preferential interactions are higher for hydrophobic residues, as Trp_{43} and Phe_{52}, relevant for neutral ones, weak for polar residues and disfavored for charged residues, especially for negatively charged ones. Predominant interactions with the protein backbone, as inferred in previous papers[55, 54] are investigated calculating the backbone contribution to the preferential coefficient (Figure 4.9).

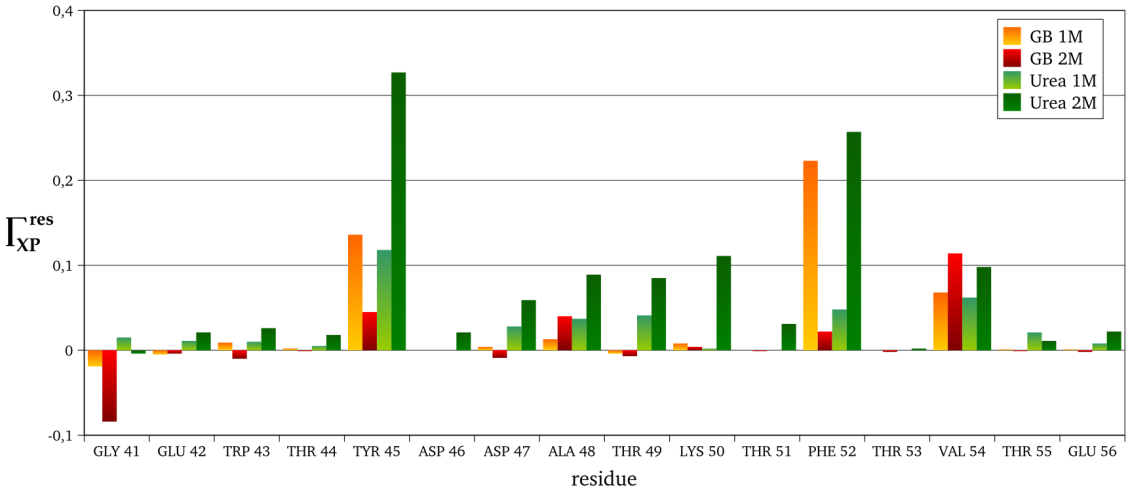

Figure 7.12: Backbone contribution to the per residue preferential coefficient. While GB values are strongly reminiscent of the overall Γ_{XP}, urea values are always positive, even for charged groups, suggesting a strong urea-backbone interaction.

Backbone contributions supported previous considerations, showing once more the exclusion of GB from the protein surface with the exception of hydrophobic residues. For urea, Γ_{XP} for the backbone were, with only one exception, always positive, even for charged residues, suggesting a strong interaction between urea and the peptidic main-chain. Once more, results for urea strongly agreed with previously reported ones[57], where it was suggested that urea acts as a "backbone surrogate". In-depth analyses of the contacts between protein and co-solvent, within a shell of 4.5 Å, confirmed once more what observed for the co-solvent distribution. Compared to water molecules, co-solvents seemed to spend most of the time around hydrophobic residues: the majority of the contacts with the protein, for both urea and GB, involved Trp_{43}, Tyr_{45} and Phe_{52} (Figure 4.10).

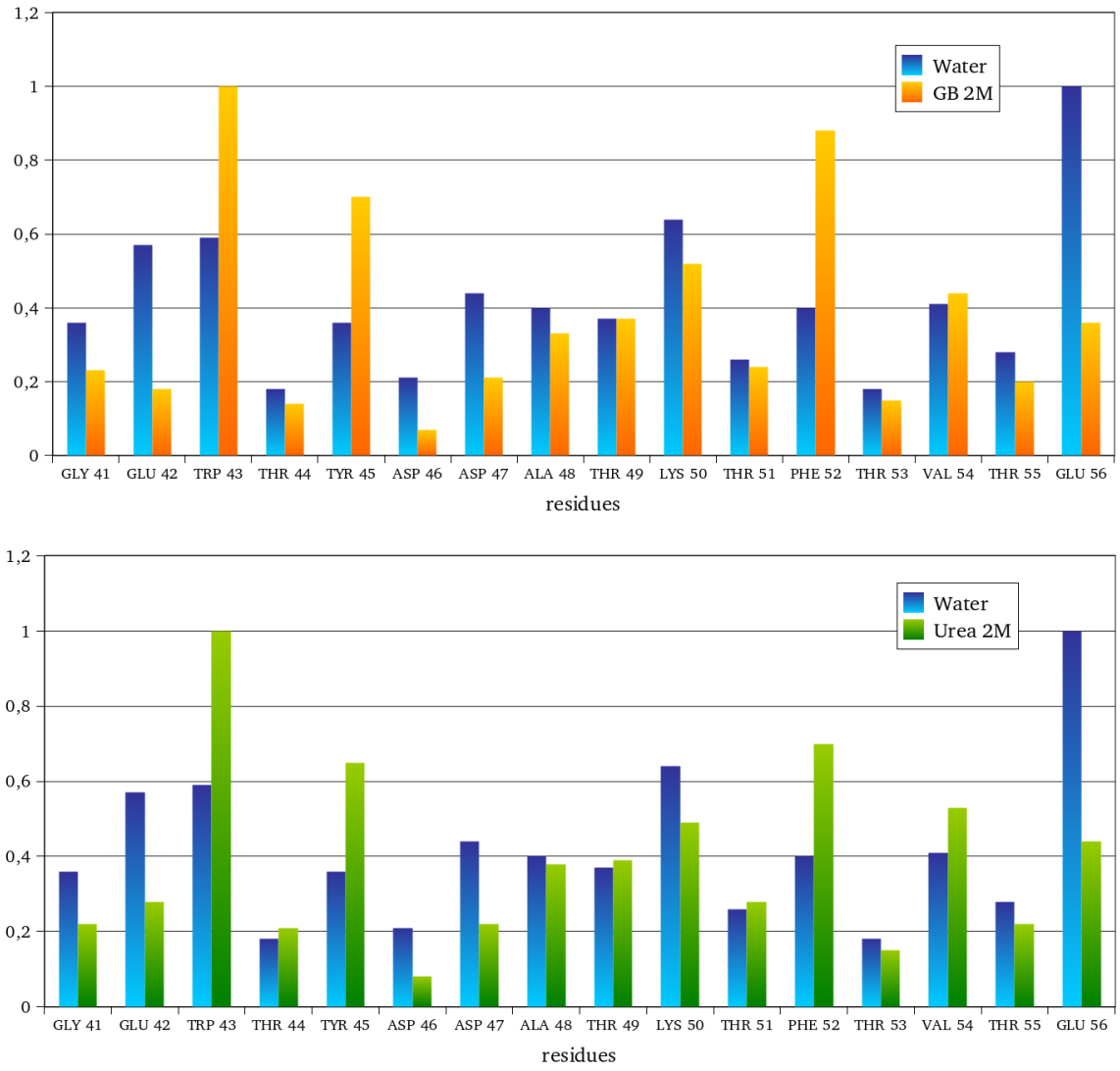

Figure 7.13: Protein/co-solvent and protein/water contacts within 4.5 Åat 350 K. (a) Results for GB. (b) Results for urea.

EXPERIMENTAL VALIDATION

NMR analyses for protein G (B1) hairpin

The following NMR tests were performed by Prof. G. Speranza, Dr. P. Francescato and Dr. T. Recca and are reported for the sake of completeness.

NOE(*)	Water (278 K)	Water (288 K)	GB (288 K)	distance (Å)	dev.st. (Å)	Urea (278 K)	distance (Å)	dev.st. (Å)
NH(K50)-NH(T49)	S	S	S	2,28	0,25	W	2,25	0,28
NH(T49)-NH(A48)	S	S	S	2,55	0,24	W	2,49	0,31
NH(T51)-NH(K50)	W	W	W	2,61	0,23	W	2,48	0,29
Hβ(D47)-Harom(Y45)	W	-	W	5,04	0,74	-	9,97	1,09
Hβ(Y45)-Harom3,5(F52)	W	-	W	4,12	1,19	-	8,78	1,97
Hβ(F52)-Harom4(W43)	W	-	W	5,64	1,48	-	11,34	3,46
Hα(Y45)- Hα(F52)	M	-	M	2,23	0,15	-	8,16	2,99
Hβ(Y45)-Harom2,4,6(F52)	W	-	W	4,36	0,97	-	8,69	2,16
Hβ(W43)-Harom2,4,6(F52)	M	-	W	7,69	1,64	-	11,5	2,49
Hα(W43)- Hα(V54)	M	-	M	2,51	0,27	-	17,04	3,3

Table 7.1: NOE signals for the peptide in water and in 1 M GB and urea.

A validation of the computational results and of the importance of the observed interaction was achieved through an accurate analyses of the NOE contacts and through their comparison with average distances calculated from the MD trajectories. The comparison revealed a strong agreement between computational and experimental results.

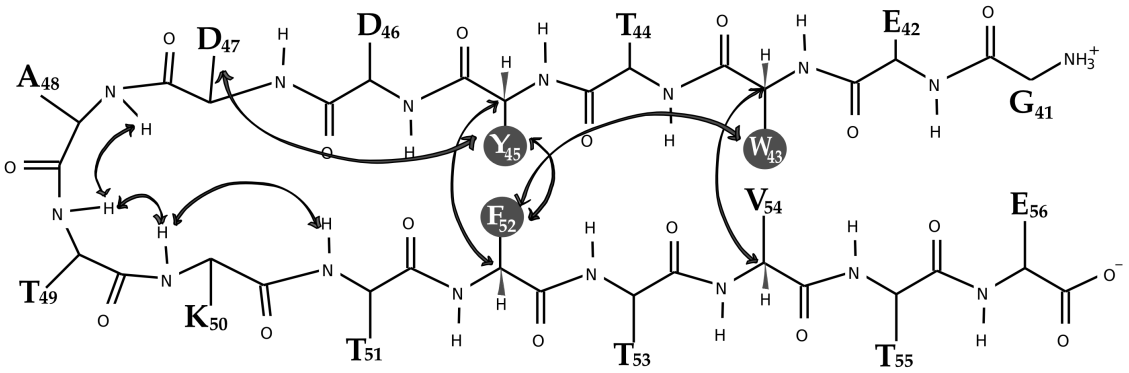

Figure 7.14: Schematic representation of the NOE contacts, typical of the β-hairpin folded conformation, between amide hydrogens in the turn region and between aromatic residues side-chains.

(*) The observed signals were ranked on the base of the average proton-proton distance as Weak (W, ca. 3.7-5.0 Å), Medium (M, ca. 2.6-3.7 Å) or Strong (S, ca. 2.2-2.6 Å). The comparison of the experimental signals and the distances calculated from the MD trajectories revealed a strong agreement, with the NMR contacts pattern correctly predicted by the simulations.

Conclusions: Introducing the "Screen-Effect"

In the whole, our approach was able to provide an excellent reproduction of the overall effect of both GB and urea, together with an insight into the order of magnitude of the stabilization/destabilization of native structures caused by the co-solvents, without a priori assumptions that some previous models were forced to introduce. Moreover, our results allowed to gain new interesting insights into osmolytes behavior and mechanism of action. The prevailing exclusion of GB from the protein surface, confirmed by Γ_{XP} calculations, strongly suggests that the protecting effect can be ascribed to an indirect effect on the bulk solvent, confirming the "indirect mechanism" hypothesis. However, our simulations also provided evidences of hydrophobic interactions between the osmolyte and some protein residues, previously strongly put forward by "direct mechanism" promoters.

In particular, the accumulation of GB around solvent-exposed hydrophobic residue might play a relevant role in the protection of this unfavorably solvated regions of the protein surface. We named this secondary effect "screen effect", since the osmolytes seem to protect hydrophobic regions from water, accumulating around them and creating a sort of "screen".

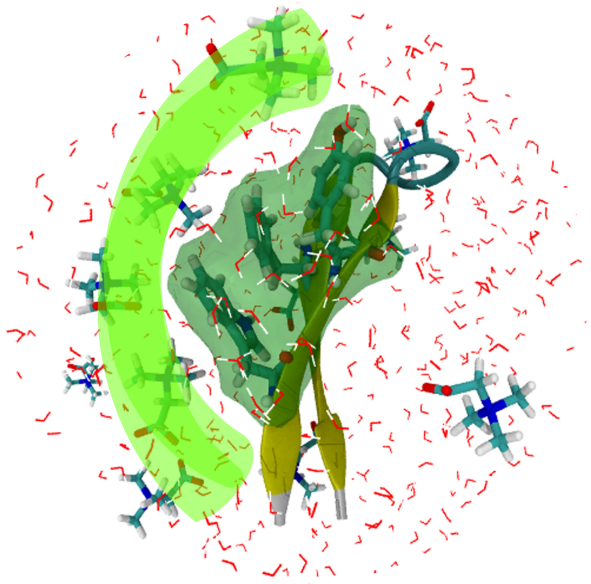

Figure 7.15: The proposed "screen effect" for protecting osmolytes. GB accumulates around hydrophobic residues relieving them from unfavorable solvation.

Considering the limited number of residues involved in these interactions, however, it is unlikely that their contribution would be the leading one. On the contrary, a

"two-fold mechanism" hypothesis could be formulated, according to which the main contribution to the protecting effect is due to the solvent, while secondary effects, highly dependent on the nature of the protein SASA, can arise from direct interactions between the osmolytes and hydrophobic residues.

CHAPTER 8

The case of a small protein

The results obtained for the $\beta-$hairpin fragment gave relevant informations about the osmolyte-mediated protecting effects. To generalize these results and further characterize these solutes behavior we performed simulations of a larger protein, the Villin headpiece helical subdomain (HP35, Figure 8.1).

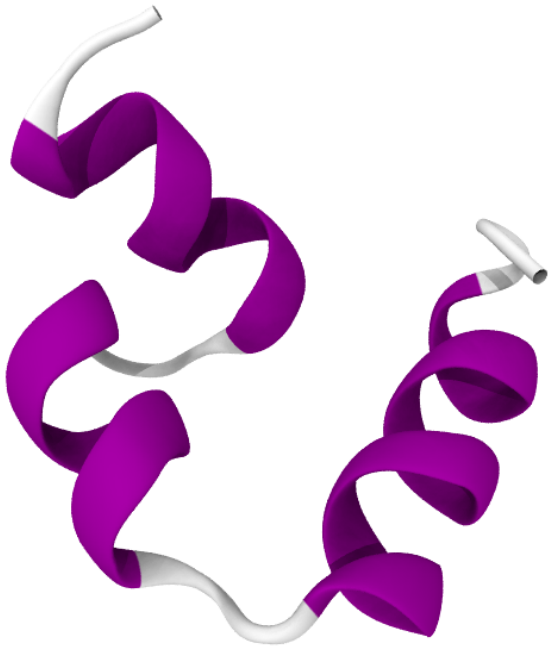

Figure 8.1: The structure of Villin headpiece helical subdomain (HP35) as reported in the PDB file.

This 35 residues C-terminal fragment of villin is one of the smallest proteins able to fold autonomously, assuming a well defined secondary and tertiary structure. HP35 is composed of three $\alpha-$helices packed together by a hydrophobic core, involving three phenylalanines (residues 7, 11 and 18). These mini-protein, because of its small size and fast dynamics, had been widely studied by means of long full-atom

MD[122][123], Replica-Exchange[124] or Monte-Carlo[125]. Recently, Piana and coworkers[126] have succeeded in correctly reproducing the dynamics of HP35 using a new approach, named Bias Exchange Molecular Dynamics (BEMD)[127]. Since for larger proteins the definition of the CVs is not trivial, we decided to follow this new approach that, as explained in the previous chapter (**Computational Methods II**), allows to use a larger set of CVs.

A good reproduction of the protein dynamics in water is crucial in order to understand the effects of osmolytes. Since accurate force-fields are essential for the success of molecular dynamics simulations, we decided to use an improved force-field, with correction by D. E. Shaw[128] and G. Hummer[129], and to carefully validate HP35 simulation in water, comparing computational results with available experimental data.

8.1 HP35 folding mechanism in pure water

The time scales that can be accessed using an unbiased MD simulation are extended everyday thanks to the constant increase in the available computational power. For this reason, we decided to start our study with a very long simulation to test which information can be obtained for HP35 using an unbiased MD. We performed a 1.5 μs long fully atomistic MD simulation at 300 K starting from the NMR structure (PDB code: 1UNC)[130]. This simulation was also performed as a test for the improved force-field used in our simulations (see **Computational Details** box). During this run we were able to observed only a few partial unfolding events, connected with a minor loss of helix 3 secondary structure.

It is worth noting that, even if the "rare event", in our case unfolding, would have been observed, the statistics connected to the simulation would have been flawed: the system would have spent most of the simulation time in the initial basin, sampling the same basin over and over again. When the "rare event"would have finally taken place, we would have obtained a trajectory, presumably a very long one, where only one unfolding event was observed. This means that, even if MD is approaching the time scales needed for large conformational transitions (at least several μs), as those involved in folding, it is still not able to provide accurate and reliable information

on the event itself, since MD can barely observe the process and a good sampling of "many rare events", i.e. a statistically meaningful ensemble of them, is still beyond its reach.

COMPUTATIONAL DETAILS

HP35 structure was obtained from the Protein Data Bank (PDB code: 1UNC). The protein was solvated with TIP3P[107] waters in a 50.0 Åcubic box and neutralized with 2 Cl^- ions. Amber99sb force field was used, including corrections by D. E. Shaw[128] and G. Hummer[129]. GROMACS 4 package[131] was used to perform the calculations. All bonds were constrained to their equilibrium distance using LINCS algorithm[132]. The system was minimized and equilibrated at constant pressure using a Berendsen barostat[133] for 1000 ps. For the standard MD, a 1.5 μs productive phase was performed under NVT ensemble, using a V-Rescale algorithm[134] for temperature coupling.

The long MD simulation was able, anyway, to give some piece of information. In particular, we observed two different folded structures that differ for the "compactness" of the hydrophobic core. The solvent accessible surface areas (SASA) for the the two conformations were 45.57 nm^2 and 48.13 nm^2 (Figure 8.2). No differences were observed concerning the secondary structure of these states. Moreover, analyses of the conformations with the highest RMSD from the initial structure revealed the presence of the aforementioned partially unfolded state in which helix 3 is largely unstructured.

These results revealed to be in agreement with high-resolution triplet-triplet relaxation experiments (TTET)[135] observing the existence of two different folded states, named N and N'. The higher stability of the region encompassing helix 1 and 2 has also been pointed out by Tang and coworkers[136] who performed temperature-dependent NMR analyses on different peptides, observing that the subsequence corresponding to the first two helices is able to fold correctly, hence even without tertiary contacts with helix 3. The relative stability of the three helices observed by TTET and NMR seems to be correctly reproduced by the recently updated force-field that we employed.

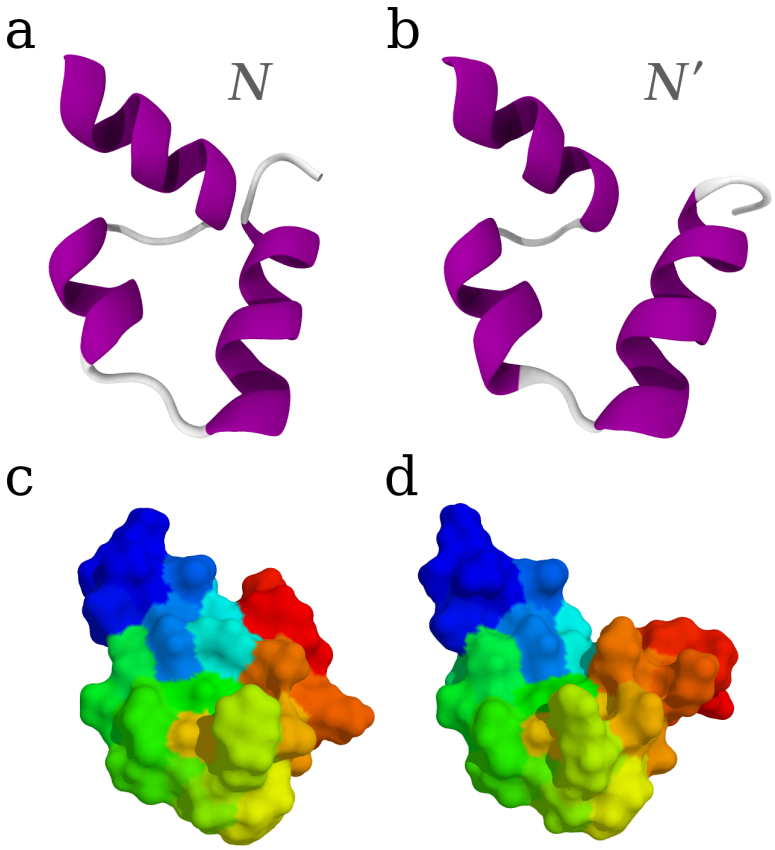

Figure 8.2: The two folded states, named N and N' according to Kiefhaber at al.[135] (a,b). The difference between N and N' can be ascribed to a more compact hydrophobic core in the N state, as the difference solvent accessible surface area confirmed (c,d).

As anticipated, to reconstruct a fully converged free-energy landscape of the HP35 folding we used massive Bias Exchange Molecular Dynamics simulations (BEMD).

COMPUTATIONAL DETAILS

For BEMD simulations, following the approach in the reference article by Piana et al.[126], 8 different replicas were run: a neutral replica on which no bias was applied (but still able to exchange conformations with the others), and 7 biased replicas, each with a potential along one collective variable. The MD parameters were the same of the previous simulation. The select CVs were the number of hydrogen bonds, hydrophobic contacts and salt bridge, the correlation of the dihedral angles (index of the secondary structure content) and the α/β similarity (index of the helical content). Both the last two CV were splitted in two to describe the first and second part of the protein.

The BEMD free energy profiles confirmed the existence of two low-free-energy folded structures, corresponding to the definition of the folded states N and N' with the latter conformation being a more open structure with increased flexibility and weakened side-chain interactions. For a better comparison of computational and experimental results, we calculated the free energy as a function of the C_α RMSD from the NMR structure and the distance between Trp_{23} and Phe_{35}, measured in the TTET experiments (Figure 8.3).

Two well defined minima corresponding to the N and N' states were observed: as opposed to the narrow N state minimum, the N' state showed a wider free energy basin, confirming its increased flexibility. As suggested by Kiefhaber et al., the N and N' conformations are very similar, reporting an overall C_α RMSD of less than 0.9 Å. Analyzing the structures of N and N', we found out that the structural differences are due to the formation of a hydrophobic core involving the last residues of helix 1 and the C-terminal residues. In particular, the stacking of the aromatic rings of Phe_{10} and Phe_{35} seemed to be responsible for the observed slightly more compact structure in which the orientation of helix 1 slenderly changes (Figure 8.4).

In a previous work Pande *et al.*[123] suggested that the stacking of Phe_{10} and Phe_{35} constitutes a kinetic trap since their strong interaction prevents HP35 from adopting a correct folding. However, it was proved experimentally[137] that mutation of residue Phe_{35} does not cause considerable alteration of folding kinetics. Considering, these evidences, we suggest that, in agreement with experimental results and with

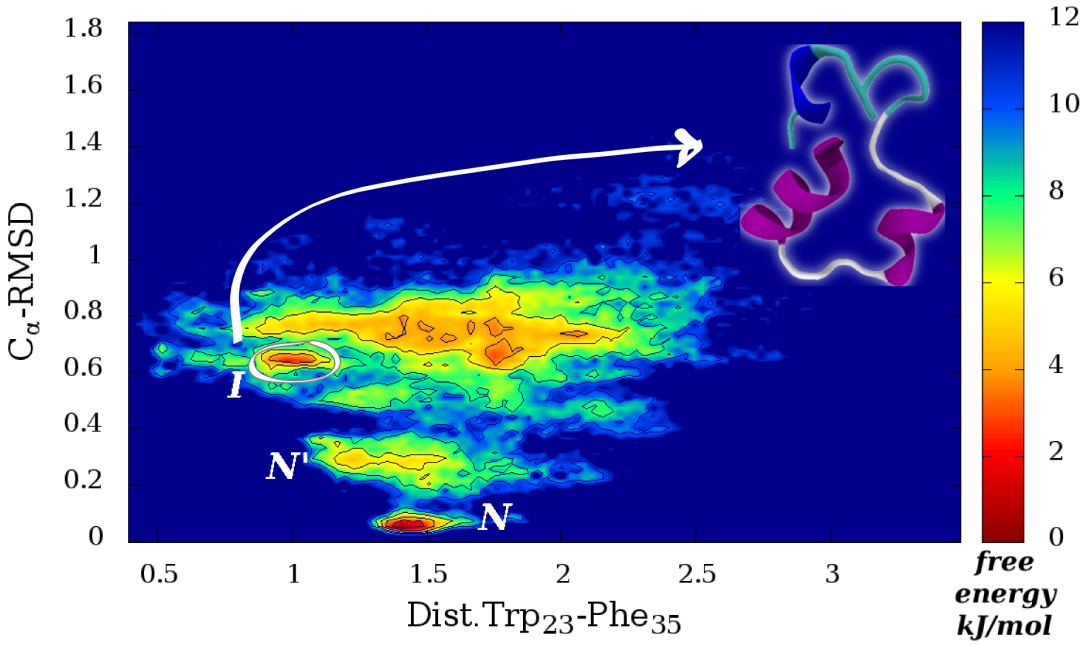

Figure 8.3: Free energy landscape as a function of the distance between Trp_{23} and Phe_{35}, the same measured in the TTET experiments. All the experimentally identified conformation can be observed.

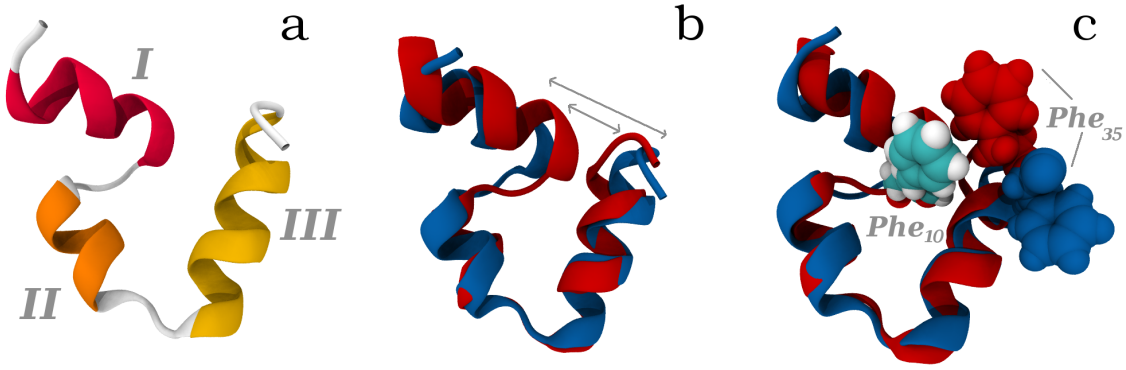

Figure 8.4: The locking mechanism, postulated by Kiefhaber et al.[135] is due to the stacking of the aromatic rings of Phe_{10} and Phe_{35}.

the proposed locking/unlocking model, the strong interaction between the two phenylalanines locks the N state once the correct folding is achieved. Besides N and N', a secondary partially unfolded minimum, corresponding to a low value of the distance, can be observed, as observed in the 1.5 μs MD. This basin, corresponding to a Trp_{23}-Phe_{35} distance of 9.85 Å, against the 15.22 Å of the native state, is populated by conformations in which helix 3 is partially unfolded. Interestingly, this conformation, belonging neither to the folded nor unfolded basins, strongly

resembles the intermediate state I proposed on the basis of experimental results[135] and is consistent with the prediction by Eaton et al.[138] that the intermediate state should lie on the folded side of the major free energy barrier and with the observed relative stability of the three helices[136, 139]. The structural features of the I state are also consistent with the well-documented observation of helical residual structure in the unfolded ensemble of HP35[136, 139–141] that, in turn, strongly agrees with the results of our simulations (see Figure 8.5 below).

8.2 HP35 in osmolytes solutions

To investigate the effects of osmolytes on the stability of HP35, we performed BEMD simulations of three systems: HP35 in 1M GB, in 1M TMAO and in 1M urea and compared the results with those obtained in pure water.

Free Energy Profiles

The free energy profiles were calculated from the statistics accumulated on the neutral (i.e. unbiased) replica. Indeed, if the run is long enough and the sampling is efficient, the probability distribution of the different states can be obtained from the occupancy of the states, monitored along the trajectory. Analyses of the profiles revealed striking differences between simulation in water and in 1M osmolytes solution (Figure 8.5). Analyses of $\alpha/\beta-$similarity revealed a larger helical content for the simulations in GB and TMAO (fig. 8.5a,b), but the most significant differences were found analysing the number of hydrophobic contacts and salt-bridges (Figure 8.5c,d).
The population of the two main minima in water, corresponding to CV values of $\sim$18 and $\sim$27 were inverted in the simulations with the osmolytes, revealing a tendency to form less hydrophobic contacts. As highlighted before, the most favored minimum in 1M GB and TMAO was, nevertheless, compact and fully structured, suggesting that the lower value was not due to the loss of any particular contact, but more likely, to an overall relaxation of the hydrophobic core. On the other hand, the profile along the number of salt-bridges revealed a secondary less pronounced minimum for the osmolytes solutions simulations, with a slightly higher value of the CV (Figure 8.5d).

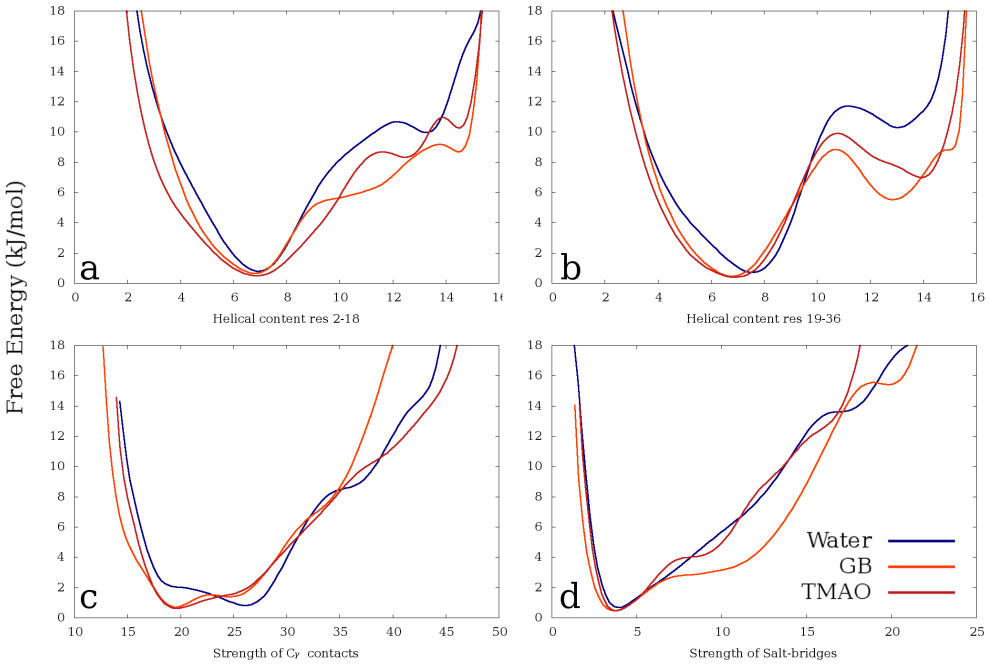

Figure 8.5: Free energy profiles as a function of the helical content (first and second part of the protein), the strength of hydrophobic contacts and the strength of salt-bridges.

Energy Landscapes and Structural Analyses

Unfortunately, we were not able to observe two well separated basins for the folded and unfolded states in the monodimensional profiles. However, the definition of a folded basin F and an unfolded basin U is mandatory to calculate the ΔG_{unfold} (see [1] on page 95). To have a more detailed picture of the different conformations, we represented the free energy along an additional degree of freedom, as to obtain a bi-dimensional map. We decided to calculate the Root Mean Square Deviation (RMSD) from the NMR structure along the trajectory and to use it as a secondary CV. The free energy surfaces, while containing a much higher degree of informations, highlighting features that were concealed in the monodimensional profiles (Figure 8.6), didn't showed a clear distinction between the folded and the unfolded basin, as observed for the monodimensional profiles. This means that none of the selected CV is able to describe the folding of HP35 alone, or coupled with the RMSD, suggesting that the use of BEMD, and hence of a higher number of CVs, was indeed appropriate. We calculated free energy surfaces as a function of several CVs and using different combinations of them. In the end, we were able to obtain a 3D representation of the

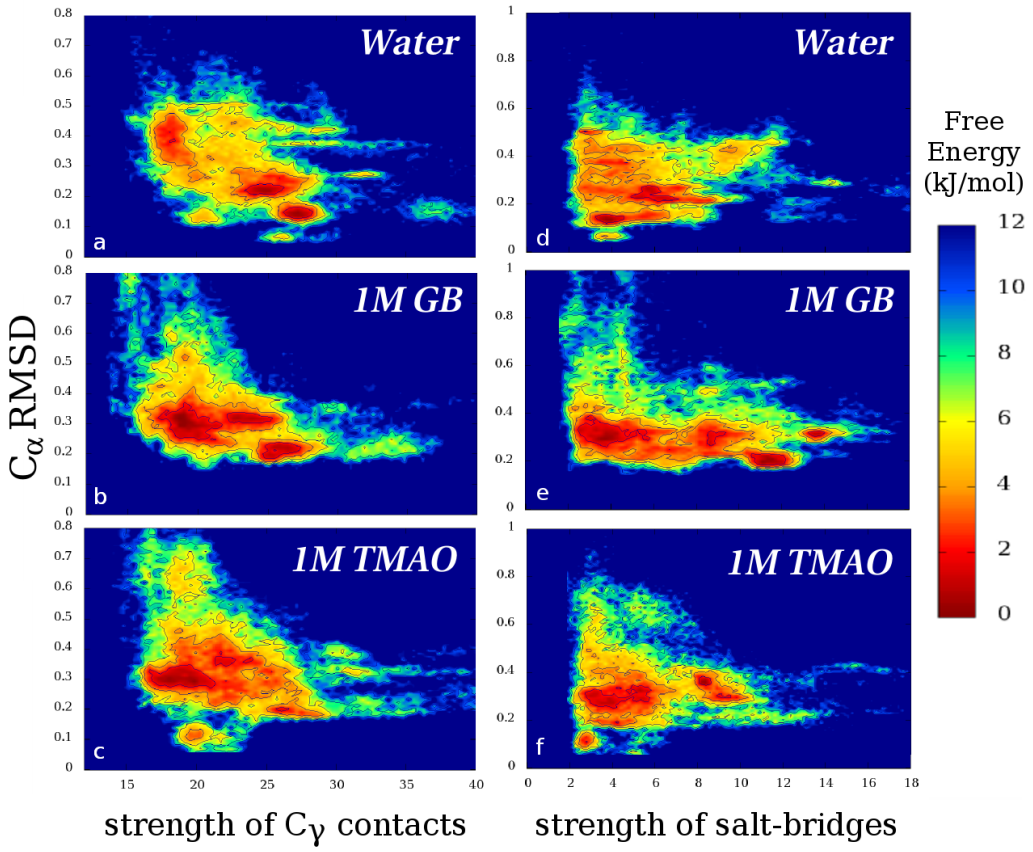

Figure 8.6: Free energy surfaces as a function of the C_α RMSD and of the strength of hydrophobic contacts (left) and the strength of salt-bridges (right).

free energy were the folded and unfolded basins can be distinguished (Figure 8.7). This is crucial for the calculation of a ΔG of unfolding. The three selected CVs were the deviation from perfect α-helix for each of the three helices taken individually. From the reported free energy surfaces, we were finally able to calculate a ΔG of unfolding (Table 8.1).

system	$\Delta G_{unfold}(\frac{kJ}{mol})$
Water	-0.17
TMAO 1M	2.89
GB 1M	2.55
Urea 1M	-2.09

Table 8.1: Free energy of unfolding as calculated from the free energy landscapes reported in Figure 8.7

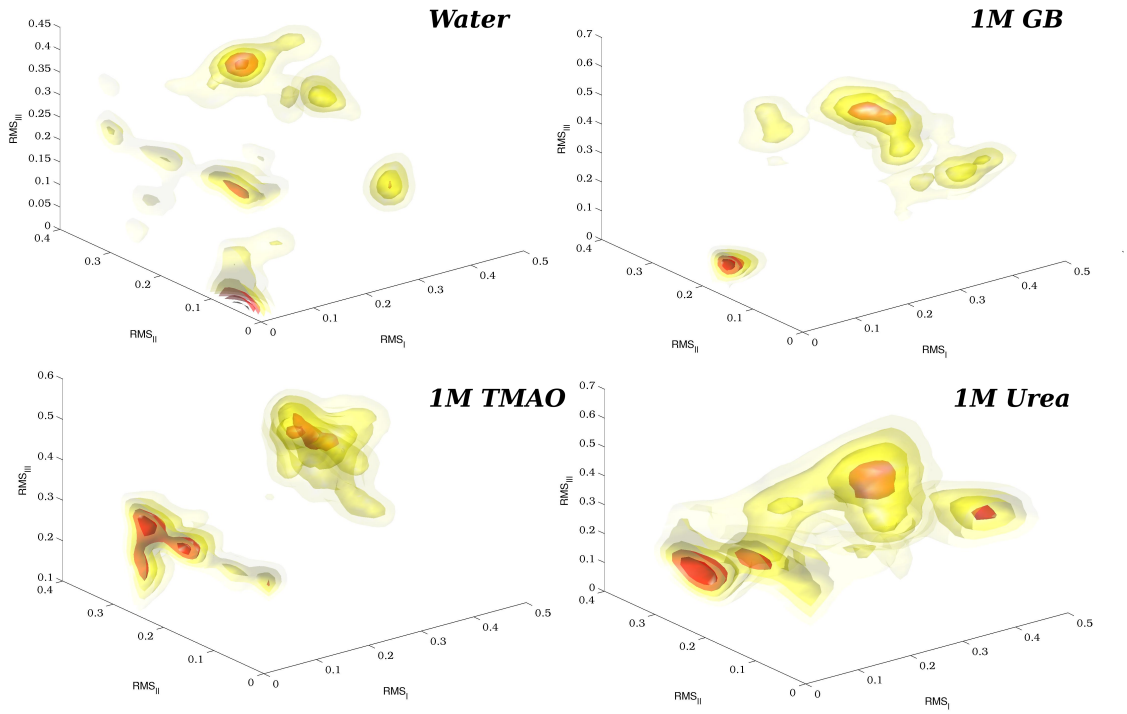

Figure 8.7: 3D free energy landscapes as a function of the deviation from perfect helix for helices I (x axis), II (y axis) and III (z axis).

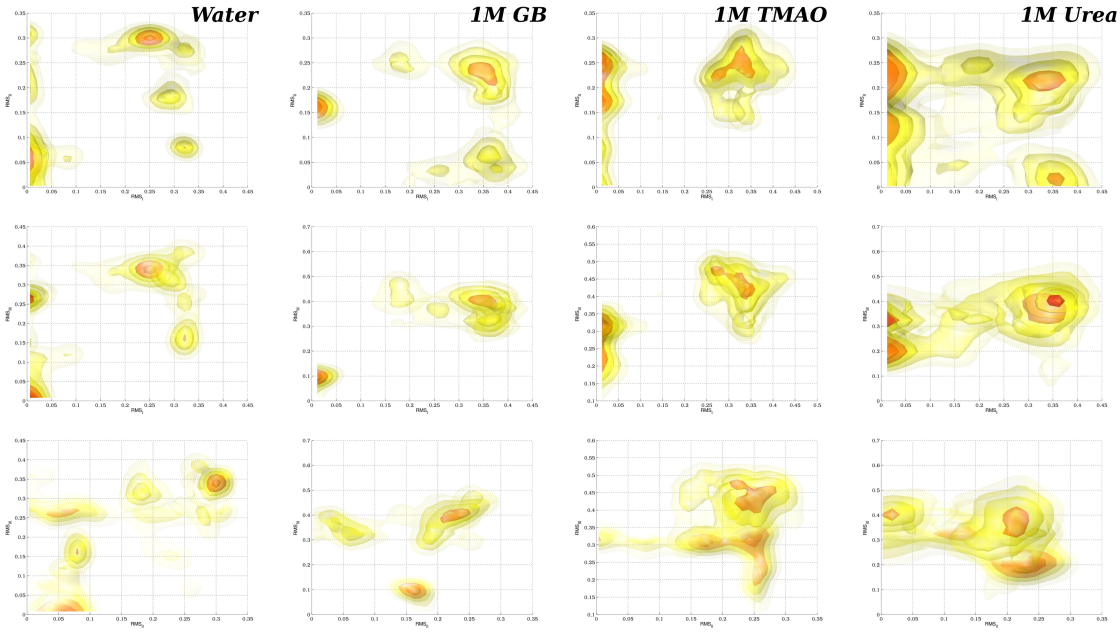

Figure 8.8: Map projection of the XY, XZ and YZ planes for all the free energy landscapes reported in Figure 8.7

The values reported in Table 8.1 confirmed that our approach was able to reproduce both the protecting effects of GB and TMAO and the denaturing effects of urea.

8.3 Mixed Solvents Analyses

Preferential coefficient and Solvent Density Function

To investigate the role of the protecting osmolytes in the stabilization of the different structures, we analyse the distribution of GB and TMAO molecules around the protein. Calculation of the Solvent Density Function (SDF) for the osmolyte molecules returned results in excellent agreement with the ones obtained in our previous study on the G protein β-hairpin fragment (previous chapter). No relevant peak was observed, with values slightly higher than unity and slowly approaching it (Figure 8.9a) for both the protecting osmolytes. The preferential coefficient Γ_{XP} was calculated to confirm the evidences of the osmophobic effect. Choosing a cut-off of 4.5 Å for the boundary between protein and bulk domains, we obtained a value of the total preferential coefficient of **-0.52** and **-1.08** for GB and TMAO, respectively, meaning that the osmolytes preferred the bulk over the protein domain. The corresponding value for HP35 in the denaturant solution was **5.22**, confirming the presence of a shell of urea molecules around the protein. Interestingly, for both protecting osmolytes and urea the most relevant contribution to Γ_{XP} came from the backbone, agreeing with previously published results founding backbone interactions as the most relevant. Γ_{XP} as a function of the threshold for protein/bulk separation is reported in Figure8.9b.

Dielectric Properties

The major differences between simulations in water and in 1 M osmolytes solutions arose when considering contacts between hydrophobic or charged residues; since the preferential coefficient revealed a lack of interactions between the osmolyte and the protein, we investigated the dielectric properties of the osmolyte solutions, in order to identify possible alteration of the solvent properties in respect to pure water. We calculated the static dielectric constant ϵ according to Neumann formulation[142], as described by van der Spoel et al.[143]; results are reported in Table 8.2.

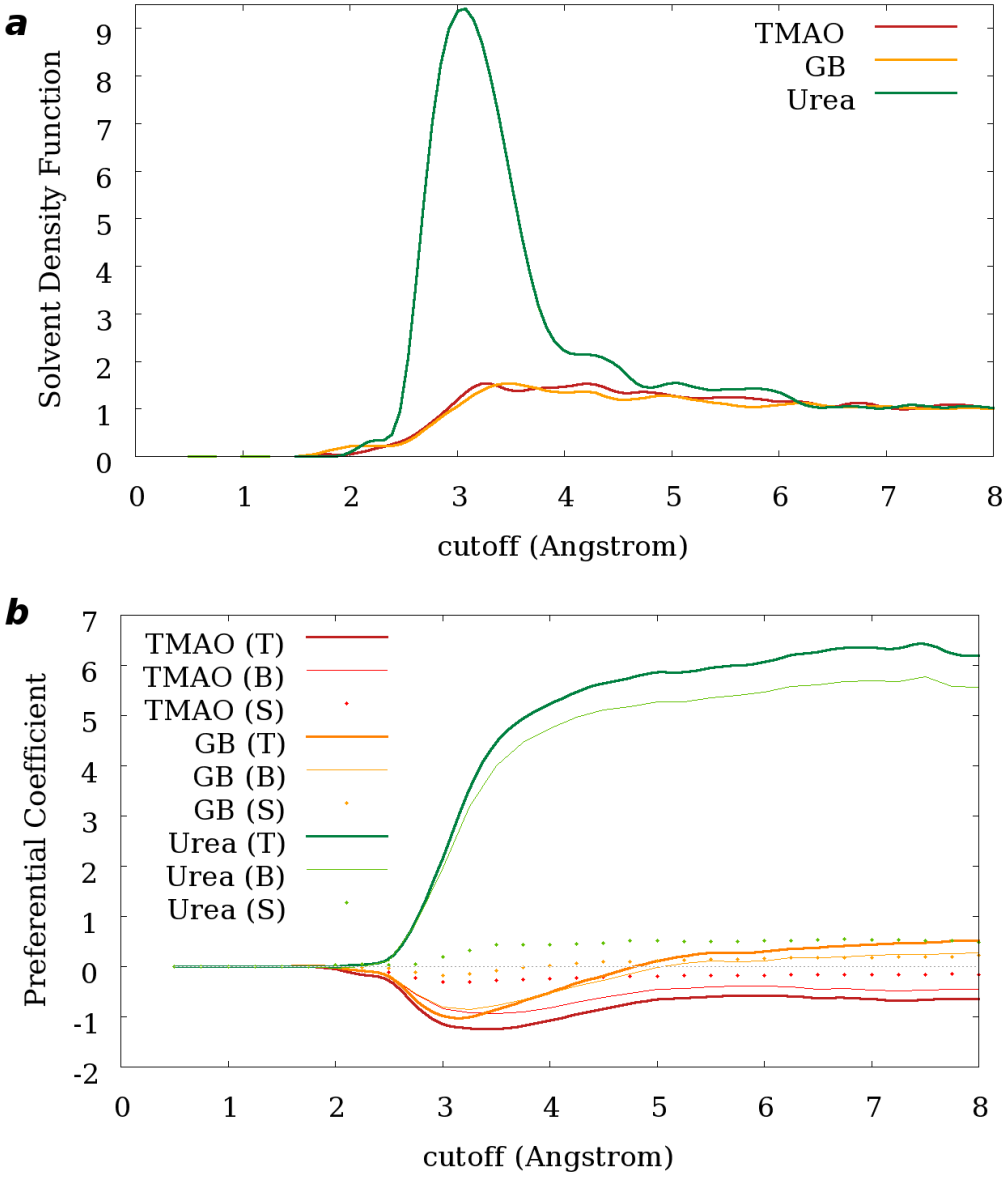

Figure 8.9: Solvent Density Function (a) and Preferential Coefficient (b) for 1M GB, 1M TMAO and 1M Urea. The osmolytes have no preferential distribution, showing no relevant peak in the SDF and reporting a negative value of Γ_{XP} in the 3-5 Åregion. Urea had the typical prominent peak around 2.8 Åand a positive value of Γ_{XP}

While a little decrease was observed for urea, both protecting osmolytes showed an increase of the overall ϵ of the solution, in agreement with experimental results[144]. However, averaging the calculation over water molecules only, a decrease of the ϵ was observed. This means that while the overall ϵ value was increased due to the strong dipole of GB and TMAO, water molecules behaved as if their dipoles were decreased. Since osmolytes are excluded from the protein surfaces, it is reasonable to assume that

system	Total ϵ	Water ϵ
Water	98.0	98.0
TMAO 1M	103.0	90.3
GB 1M	106.4	83.0
Urea 1M	95.2	92.5

Table 8.2: Dielectric constant for water and the different 1M solutions. Values obtained averaging over water molecules only are reported in the second column.

the main effect perceived by the protein itself might be the one connected with water decreased ϵ. Interestingly, most of the previously reported features of osmolytes behavior (e.g. folded state protection, osmophobic effect and backbone repulsion) could be solidly explained in the context of a reduced polarity of the watery solvent. To confirm these hypotheses, we performed a BEMD simulation of HP35 in modified TIP3P water molecules (W_{79}), whose charges had been altered to reproduce a 20% decrease of the dielectric constant. The free energy profiles obtained reported a striking similarity with the ones obtained for the osmolyte solutions (Figure 8.10).

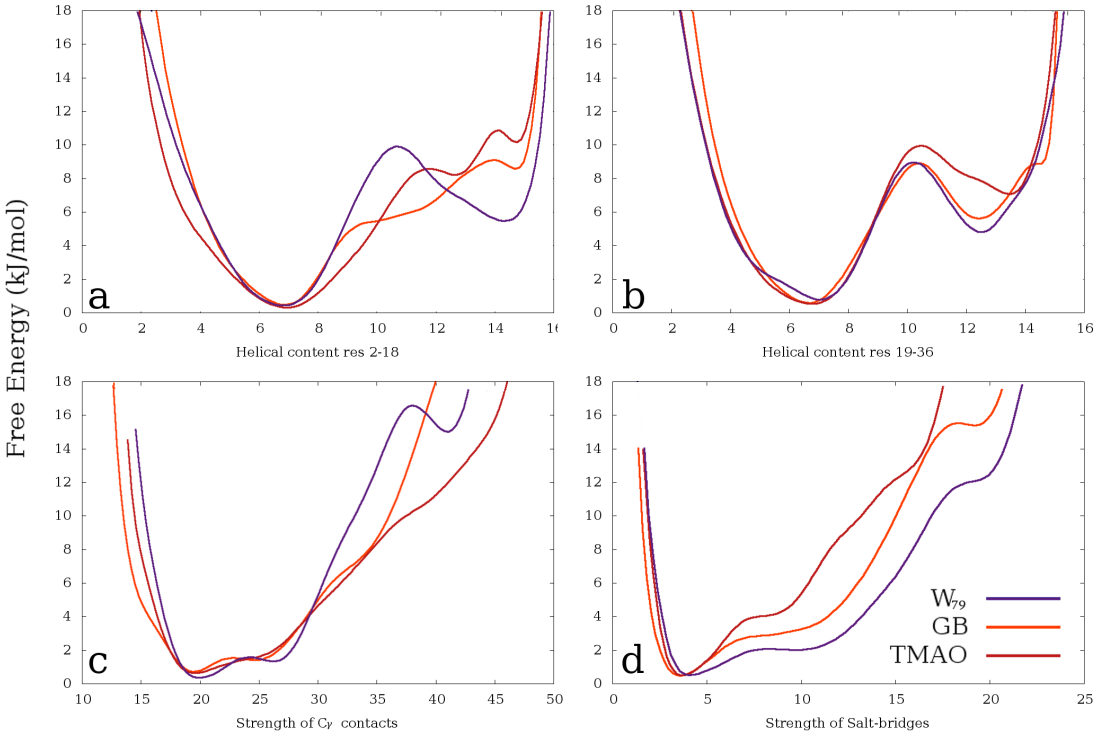

Figure 8.10: cap.

Analogies could be found also on a structural level, with the folded states in W_{79}

and the osmolytes presenting the same salt-bridge not observed for the simulation in pure water (Figure 8.11). In the whole, simulation in W_{79} waters revealed that the change in the dielectric constant reproduced properly most of the features of the osmolytes solutions free energy landscape. While minor effects might still be uncovered, a relevant role could be ascribed to the ϵ shifting, reasonably identifiable as the foremost effect caused by the osmolyte and as the main driving force beneath native state protection.

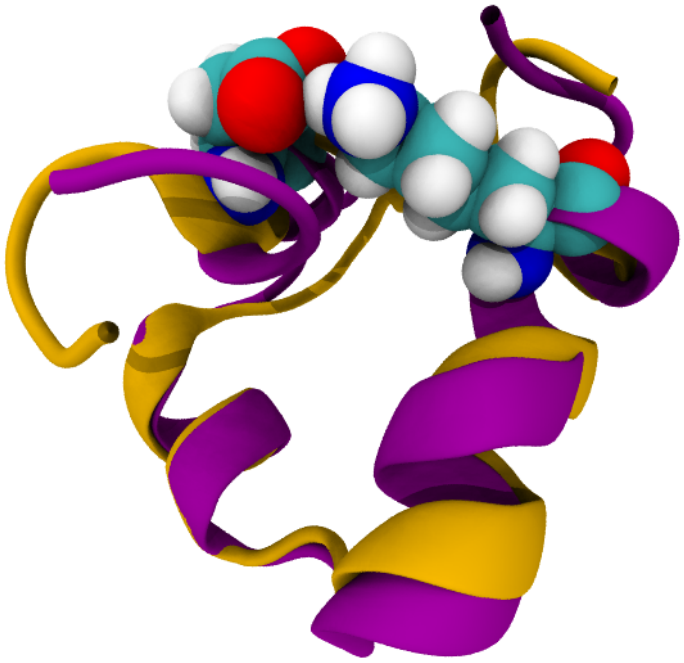

Figure 8.11: The strong salt-bridge observed in 1M osmolytes solutions and in the W_{79} simulation, but not in pure water.

EXPERIMENTAL VALIDATION

Experimental analyses of HP35

The following biological tests were performed by Dr. M. Marenchino and are reported for the sake of completeness.

Thermal and chemical denaturation were exploited to gain a in-depth thermodynamic description of the stabilization effects due to the osmolytes. Equilibrium thermal unfolding measurements were performed on HP35 in water and in 1M TMAO or GB solutions (Figure 8.12). In addition, the stability of HP35 was investigated in a 0.66 M NaCl solution. This salt concentration reduces the dielectric of the solvent to 67.2 (water experimental value is 78.0), thus simulating the decrease of ϵ due to the osmolytes. From the combination of the change in Gibbs free energy upon thermal and chemical denaturation, the stability plot of HP35 was obtained.

	$\Delta G^{exp}_{unfold}(\frac{kJ}{mol})$	$\Delta G^{calc}_{unfold}(\frac{kJ}{mol})$	T_m (^{o}C)
Water	-0.17	-1.00	44
TMAO 1M	2.89	1.63	51
GB 1M	2.55	0.80	49
NaCl 0.66M	2.89	$0.67^{(\ddagger)}$	49

Table 8.3: Thermodynamic data obtained from equilibrium thermal unfolding measurements.

The data reported in Figure 8.12 and Table 8.3 confirmed computational results. The values for NaCl 0.66, in particular, confirmed that a decrease of the dielectric constant is consistent with a protecting effect of the same magnitude of the one observed for the osmolytes.

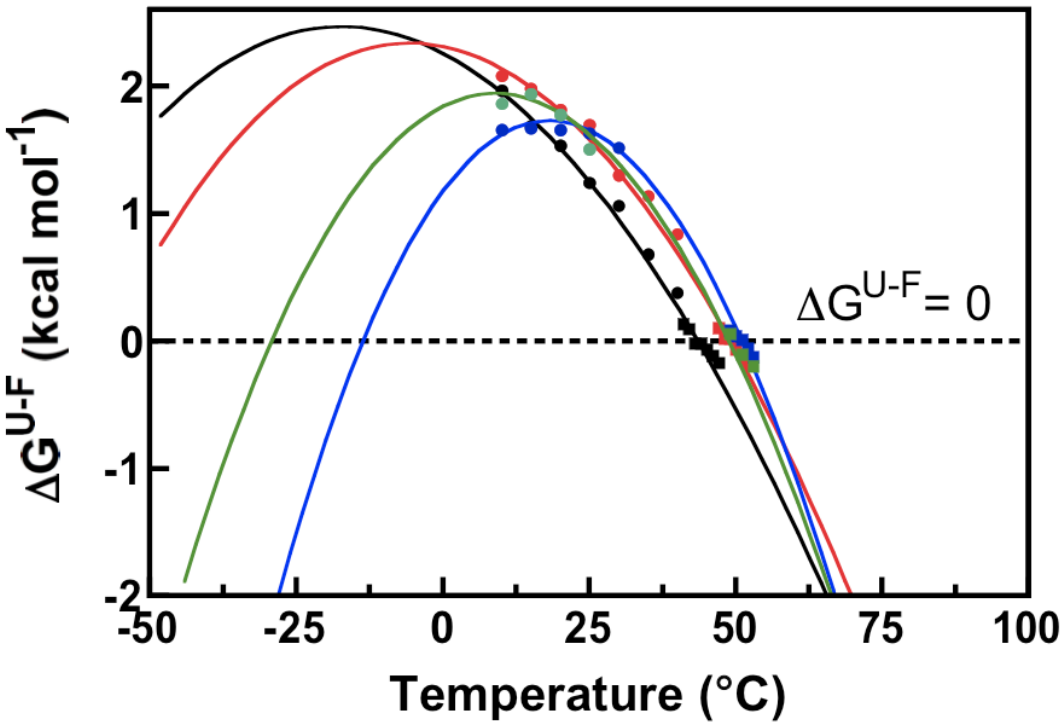

Figure 8.12: Protein stability curves for HP35 in water (black), 0.66 M NaCl (red), 1 M GB (green), or 1M TMAO (blue). Squares represent unfolding free energies measured directly from the transition zones of the thermal denaturation curves shown in Figures 2 at HP 35 concentration of 50 μM. Circles represent $\Delta G_U^{H_2O}$ values determined from an analysis of urea denaturation curves determined at various temperatures. Solid lines show the best fitting to the Gibbs-Helmholtz equation.

Conclusions

The pursuit of a universal framework for the explanation of osmolytes effects have drawn considerable attention in the last decades, due to the wide range of possible application that a deeper understanding of these molecules behavior would unleash, starting from, but not limited to, crops resistance enhancements for regions with severe climatic conditions. However, despite multiple efforts and an ever increasing interest, a satisfactory explanation of the mechanism beneath osmolyte-mediated proteins protection and cells adaptability remains an unfulfilled issue. Our studies, focused on the effects of osmolytes, both protecting and denaturing ones, on a small β-hairpin fragment and on HP35 revealed a complex scenario for osmolytes mechanism of action. The scatteredly reported evidences of both direct and indirect effects were successfully explained in the setting of a "double mechanism". Exclusion from the protein surface, largely confirmed by our simulation, suggests a prevailing indirect mechanism, but secondary interactions with hydrophobic residues are constantly observed, contributing to what we've termed the "screen effect". In addition to the important understanding of direct-indirect effects interplay, our studies also strongly suggested that the main driving force behind the protecting effect might be a "ϵ-shifting" of the solution, and in particular a change of the dielectric properties of the watery solvent. While an unarguable demonstration is still to be achieved, our approach have succeeded in sheding more light on osmolytes effects and in unleashing new interesting ideas.

References for Part II

[1] Dill, K. A. (1990) Dominant forces in protein folding. *Biochemistry 29*, 7133–7155.

[2] Fersht, A. (1997) Nucleation mechanisms in protein folding. *Curr Opin Struct Biol 7*, 3–9.

[3] Yancey, P. H., Clark, M. E., Hand, S. C., Bowlus, R. D., and Somero, G. N. (1982) Living with water stress: evolution of osmolyte systems. *Science 217*, 1214.

[4] Arakawa, T. and Timasheff, S. N. (1985) The stabilization of proteins by osmolytes. *Biophys J 47*, 411–414.

[5] Kinne, R. K. H. (1993) The role of organic osmolytes in osmoregulation: from bacteria to mammals. *J Exp Zool 265*, 346–355.

[6] Yancey, P. H. (2005) Commentary Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *J Exp Biol* pp. 2819–2830.

[7] Burg, M. B. and Ferraris, J. D. (2008) Intracellular organic osmolytes: function and regulation. *J Biol Chem 283*, 7309–7313.

[8] Yancey, P. H. and Burg, M. B. (1989) Distribution of major organic osmolytes in rabbit kidneys in diuresis and antidiuresis. *Am J Physiol 257*, 602.

[9] Weber, D. J. (2009) *Adaptive Mechanisms of Halophytes in Desert Regions.* (Springer, Netherlands).

[10] Yancey, P. H., Rhea, M. D., Kemp, K. M., and Bailey, D. M. (2004) Trimethylamine oxide, betaine and other osmolytes in deep-sea animals: depth trends and effects on enzymes under hydrostatic pressure. *Cell Mol Bio 50*, 371.

[11] Gillett, M. B., Suko, J. R., Santoso, F. O., and Yancey, P. H. (1997) Elevated levels of trimethylamine oxide in muscles of deep-sea gadiform teleosts: A high-pressure adaptation? *J Exp Zool 279*, 386–391.

[12] Storey, K. B. (1997) Organic solutes in freezing tolerance. *Comp Biochem Physiol 117*, 319–326.

[13] Yayanos, A. A., Dietz, A. S., and Boxtel, R. V. (1979) Isolation of a deep-sea barophilic bacterium and some of its growth characteristics. *Science 205*, 808.

[14] Nogi, Y., Kato, C., and Horikoshi, K. (2002) Psychromonas kaikoae sp. nov., a novel piezophilic bacterium from the deepest cold-seep sediments in the Japan Trench. *Int J Syst Evol Microbiol 52*, 1527.

[15] Yayanos, A. A. (1986) Evolutional and ecological implications of the properties of deep-sea barophilic bacteria. *Proc Natl Acad Sci USA 83*, 9542.

[16] Takai, K., Nakamura, K., Toki, T., Tsunogai, U., Miyazaki, M., Miyazaki, J., Hirayama, H., Nakagawa, S., Nunoura, T., and Horikoshi, K. (2008) Cell proliferation at 122 C and isotopically heavy CH4 production by a hyperthermophilic methanogen under high-pressure cultivation. *Proc Natl Acad Sci USA 105*, 10949.

[17] Brochier, C., Forterre, P., and Gribaldo, S. (2004) Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the Methanopyrus kandleri paradox. *Genome Biol 5*, 17.

[18] Furini, A., Koncz, C., Salamini, F., and Bartels, D. (1997) High level transcription of a member of a repeated gene family confers dehydration tolerance to callus tissue of Craterostigma plantagineum. *EMBO J 16*, 3599–3608.

[19] Deng, X., Phillips, J., Meijer, A. H., Salamini, F., and Bartels, D. (2002) Characterization of five novel dehydration-responsive homeodomain leucine zipper genes from the resurrection plant Craterostigma plantagineum. *Plant Mol Biol 49*, 601–610.

[20] Baldwin, W. W., Myer, R., Kung, T., Anderson, E., and Koch, A. L. (1995) Growth and buoyant density of Escherichia coli at very low osmolarities. *J Bacteriol 177*, 235.

[21] Cayley, S., Lewis, B. A., Guttman, H. J., Record, M. T., and Others. (1991) Characterization of the cytoplasm of Escherichia coli K-12 as a function of external

osmolarity* 1:: Implications for protein-DNA interactions in vivo. *J Mol Biol 222*, 281–300.

[22] Csonka, L. N. (1989) Physiological and genetic responses of bacteria to osmotic stress. *Microbiol Rev 53*, 121.

[23] Lai, M. C., Yang, D. R., and Chuang, M. J. (1999) Regulatory factors associated with synthesis of the osmolyte glycine betaine in the halophilic methanoarchaeon Methanohalophilus portucalensis. *App Env Microbiol 65*, 828.

[24] Singer, M. A. and Lindquist, S. (1998) Thermotolerance in Saccharomyces cerevisiae: the yin and yang of trehalose. *Trends Biotechnol 16*, 460–468.

[25] Qu, Y., Bolen, C. L., and Bolen, D. W. (1998) Osmolyte-driven contraction of a random coil protein. *Proc Nat Acad Sci USA 95*, 9268–9273.

[26] Santoro, M. M., Liu, Y., Khan, S. M., Hou, L. X., and Bolen, D. W. (1992) Increased thermal stability of proteins in the presence of naturally occurring osmolytes. *Biochemistry 31*, 5278.

[27] Baskakov, I. V. and Bolen, D. W. (1998) Forcing thermodynamically unfolded proteins to fold. *J Biol Chem 273*, 4831.

[28] Rohr, J. M., Truong, S., Hong, T., and Yancey, P. H. (1999) Effects of ascorbic acid, aminoguanidine, sorbinil and zopolrestat on sorbitol and betaine contents in cultured rat renal cells. *Exp Biol Online 4*, 1–38.

[29] Moriyama, T., Garcia-Perez, A., Olson, A. D., and Burg, M. B. (1991) Intracellular betaine substitutes for sorbitol in protecting renal medullary cells from hypertonicity. *Am J Physiol 260*, 494.

[30] Yancey, P. H. (2001) Water Stress, Osmolytes and Proteins. *Int Comp Biol 41*, 699–709.

[31] Parida, A. K. and Das, A. B. (2005) Salt tolerance and salinity effects on plants: a review. *Eco Env Saf 60*, 324–349.

[32] Hodes, D. and Hanson, A. D. (1993) Quaternary Ammonium and Tertiary Sulfonium Compounds in Higher Plants. *Annu Rev Plant Physiol Plant Mol Biol 44*, 357.

[33] Carr, W. E. S., Netherton JC, I. I. I., Gleeson, R. A., and Derby, C. D. (1996) Stimulants of Feeding Behavior in Fish: Analyses of Tissues of Diverse Marine Organisms. *Biol Bull 190*, 149–160.

[34] Yancey, P. H., Blake, W. R., and Conley, J. (2002) Unusual organic osmolytes in deep-sea animals: adaptations to hydrostatic pressure and other perturbants* 1. *Comp Biochem Physiol A Mol Int Phys 133*, 667–676.

[35] Burg, M. B., Ferraris, J. D., and Dmitrieva, N. I. (2007) Cellular Response to Hyperosmotic Stresses. *Physiol Rev 87*, 1441–1474.

[36] Garcia-Perez, A. and Burg, M. B. (1991) Renal medullary organic osmolytes. *Physiol Rev 71*, 1081–1115.

[37] Cushman, J. C. (2001) Osmoregulation in Plants: Implications for Agriculture. *Int Comp Biol 41*, 758–769.

[38] Welch, W. J. and Brown, C. R. (1996) Influence of molecular and chemical chaperones on protein folding. *Cell Stress Ch 1*, 109–115.

[39] Howard, M., Fischer, H., Roux, J., Santos, B. C., Gullans, S. R., Yancey, P. H., and Welch, W. J. (2003) Mammalian osmolytes and S-nitrosoglutathione promote Delta F508 cystic fibrosis transmembrane conductance regulator {(CFTR)} protein maturation and function. *J Biol Chem 278*, 35159–35167.

[40] Bennion, B. J., DeMarco, M. L., and Daggett, V. (2004) Preventing misfolding of the prion protein by trimethylamine N-oxide. *Biochemistry 43*, 12955.

[41] Dai, A. (2010) Drought under global warming: a review. *Wiley Interdisciplinary Reviews: Climate Change* pp. n/a–n/a.

[42] Flowers, T. J. and Yeo, A. R. (1995) Breeding for Salinity Resistance in Crop Plants: Where Next? *Aus J Plant Physiol 22*, 875.

[43] Flowers, T. J. (2004) Improving crop salt tolerance. *J Exp Bot 55*, 307–319.

[44] Rathinasabapathi, B., Sigua, C., Ho, J., and Gage, D. A. (2000) Osmoprotectant beta-alanine betaine synthesis in the Plumbaginaceae: S-adenosyl- l-methionine dependent N-methylation of beta-alanine to its betaine is via N-methyl and N,N-dimethyl beta-alanines. *Physiologia Plantarum 109*, 225–231.

[45] Athawale, M. V., Dordick, J. S., and Garde, S. (2005) Osmolyte trimethylamine-N-oxide does not affect the strength of hydrophobic interactions: origin of osmolyte compatibility. *Biophys J 89*, 858–866.

[46] Möglich, A., Krieger, F., and Kiefhaber, T. (2005) Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains. *J Mol Biol 345*, 153–162.

[47] Zou, Q., Bennion, B. J., Daggett, V., and Murphy, K. P. (2002) The molecular mechanism of stabilization of proteins by {TMAO} and its ability to counteract the effects of urea. *J Am Chem Soc 124*, 1192–1202.

[48] Bennion, B. J. and Daggett, V. (2003) The molecular basis for the chemical denaturation of proteins by urea. *Proc Natl Acad Sci USA 100*, 5142.

[49] Timasheff, S. N. (1993) The Control of Protein Stability and Association by Weak Interactions with Water: How Do Solvents Affect These Processes? *Annu Rev Biophys Biomol Struct 22*, 67–97.

[50] Cannon, J. G., Anderson, C. F., and Record, M. T. (2007) Urea-Amide Preferential Interactions in Water: Quantitative Comparison of Model Compound Data with Biopolymer Results Using Water Accessible Surface Areas. *J Phys Chem B 111*, 9675–9685.

[51] Sharp, K. A., Madan, B., Manas, E., and Vanderkooi, J. M. (2001) Water structure changes induced by hydrophobic and polar solutes revealed by simulations and infrared spectroscopy. *J Chem Phys 114*, 1791.

[52] Feng, J., Li, X., Pei, F., Chen, X., Li, S., and Nie, Y. (2002) 1H NMR analysis for metabolites in serum and urine from rats administrated chronically with La(NO3)3. *Anal Biochem 301*, 1–7.

[53] Xie, T., Liu, D., Feng, Y., Shan, L., and Wang, J. (2007) Folding stability and cooperativity of the three forms of 1-110 residues fragment of staphylococcal nuclease. *Biophys J 92*, 2090–2107.

[54] Auton, M., Holthauzen, L. M. F., and Bolen, D. W. (2007) Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc Natl Acad Sci USA 104*, 15317.

[55] Stumpe, M. C. and Grubmuller, H. (2007) Interaction of urea with amino acids: implications for urea-induced protein denaturation. *J Am Chem Soc 129*, 16126–16131.

[56] Stumpe, M. C. and Grubmüller, H. (2008) Polar or apolar—the role of polarity for urea-induced protein denaturation. *PLoS Comput Biol 4*.

[57] Hua, L., Zhou, R., Thirumalai, D., and Berne, B. J. (2008) Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc Natl Acad Sci USA 105*, 16928.

[58] Tirado-Rives, J., Orozco, M., and Jorgensen, W. L. (1997) Molecular dynamics simulations of the unfolding of barnase in water and 8 M aqueous urea. *Biochemistry 36*, 7313–7329.

[59] O'Brien, E. P., Dima, R. I., Brooks, B., and Thirumalai, D. (2007) Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: lessons for protein denaturation mechanism. *J Am Chem Soc 129*, 7346–7353.

[60] Camilloni, C., Rocco, A. G., Eberini, I., Gianazza, E., Broglia, R. A., and Tiana, G. (2008) Urea and guanidinium chloride denature protein L in different ways in molecular dynamics simulations. *Biophys J 94*, 4654–4661.

[61] Athawale, M. V., Sarupria, S., and Garde, S. (2008) Enthalpy-entropy contributions to salt and osmolyte effects on molecular-scale hydrophobic hydration and interactions. *J Phys Chem B 112*, 5661–5670.

[62] Athawale, M. V., Dordick, J. S., and Garde, S. (2005) Osmolyte trimethylamine-N-oxide does not affect the strength of hydrophobic interactions: origin of osmolyte compatibility. *Biophys J 89*, 858–866.

[63] Lee, M.-E. and van Der Vegt, N. F. A. (2006) Does urea denature hydrophobic interactions? *J Am Chem Soc 128*, 4948–4949.

[64] Bolen, D. W. and Rose, G. D. (2008) Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu Rev Biochem 77*, 339–362.

[65] Mountain, R. D. and Thirumalai, D. (2003) Molecular dynamics simulations of end-to-end contact formation in hydrocarbon chains in water and aqueous urea solution. *J Am Chem Soc 125*, 1950–1957.

[66] Bruccoleri, R. E. and Karplus, M. (1990) Biopolymers, 29, 1847–1862. *ISI 29*.

[67] Garel, T., Niel, J. C., Orland, H., Smith, J., and Velikson, B. (1991) Comparison of a New Monte Carlo peptide conformational search procedure with high temperature molecular dynamics. *J Chem Phys 88*, 2479–2488.

[68] Luo, Z., Butcher, D. J., and Huang, Z. (1997) Molecular modeling of interleukin-8 receptor β and analysis of the receptor–ligand. *Interaction Protein Eng 10*, 10.

[69] Grubmuller, H. (1995) Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys Rev E 52*, 2893–2906.

[70] Crippen, G. M. and Scheraga, H. A. (1969) Minimization of polypeptide energy, {VIII.} Application of the deflation technique to a dipeptide. *Proc Natl Acad Sci USA 64*, 42.

[71] Schulze, B. G., Grubmuller, H., Evanseck, J. D., and Others. (2000) Functional significance of hierarchical tiers in carbonmonoxy myoglobin: conformational substates and transitions studied by conformational flooding simulations. *J Am Chem Soc 122*, 8700–8711.

[72] Torrie, G. M. and Valleau, J. P. (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J Comp Phys 23*, 187–199.

[73] Huber, T., Torda, A. E., and Gunsteren, W. F. (1994) Local elevation: a method for improving the searching properties of molecular dynamics simulation. *J Comp Aid Mol Des 8*, 695–708.

[74] Hart, R. K., Pappu, R. V., and Ponder, J. W. (2000) Exploring the similarities between potential smoothing and simulated annealing. *J Comput Chem 21*, 531–552.

[75] Rahman, J. A. and Tully, J. C. (2002) Puddle-skimming: An efficient sampling of multidimensional configuration space. *J Chem Phys 116*, 8750.

[76] Voter, A. F. (1997) A method for accelerating the molecular dynamics simulation of infrequent events. *J Chem Phys 106*, 4665–4677.

[77] Hamelberg, D., Mongan, J., and McCammon, J. A. (2004) Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J Chem Phys 120*, 11919.

[78] Piela, L., Kostrowicki, J., and Scheraga, H. A. (1989) On the multiple-minima problem in the conformational analysis of molecules: deformation of the potential energy hypersurface by the diffusion equation method. *J Phys Chem 93*, 3339–3346.

[79] Simmerling, C., Miller, J. L., and Kollman, P. A. (1998) Combined locally enhanced sampling and particle mesh Ewald as a strategy to locate the experimental structure of a nonhelical nucleic acid. *J Am Chem Soc 120*, 7149–7155.

[80] Mitsutake, A., Sugita, Y., and Okamoto, Y. (2001) Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers 60*, 96–123.

[81] Sanbonmatsu, K. Y. and Garcia, A. E. (2002) Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Prot Struc Func Gen 46*, 225–234.

[82] Frantz, D. D., Freeman, D. L., and Doll, J. D. (1990) Reducing quasi-ergodic behavior in Monte Carlo simulations by J-walking: Applications to atomic clusters. *J Chem Phys 93*, 2769.

[83] Hansmann, U. H. E. (1997) Parallel tempering algorithm for conformational studies of biological molecules. *Chem Phys Lett 281*, 140–150.

[84] Wu, X. and Wang, S. (1998) Self-guided molecular dynamics simulation for efficient conformational search. *J Phys Chem B 102*, 7238–7250.

[85] Faradjian, A. K. and Elber, R. (2004) Computing time scales from reaction coordinates by milestoning. *J Chem Phys 120*, 10880.

[86] Krivov, S. V., Chekmarev, S. F., and Karplus, M. (2002) Potential energy surfaces and conformational transitions in biomolecules: a successive confinement approach applied to a solvated tetrapeptide. *Phys Rev Lett 88*, 38101.

[87] Andricioaei, I. and Straub, J. E. (1997) On Monte Carlo and molecular dynamics methods inspired by Tsallis statistics: Methodology, optimization, and application to atomic clusters. *J Chem Phys 107*, 9117.

[88] Keasar, C., Elber, R., and Skolnick, J. (1997) Simultaneous and coupled energy optimization of homologous proteins: a new tool for structure prediction. *Folding and Design 2*, 247–259.

[89] Shirts, M. R. and Pande, V. S. (2001) Mathematical analysis of coupled parallel simulations. *Phys Rev Lett 86*, 4983–4987.

[90] Snow, C. D., Nguyen, H., Pande, V. S., and Gruebele, M. (2002) Absolute comparison of simulated and experimental protein-folding dynamics. *Nature 420*, 102–106.

[91] Zhou, R., Berne, B. J., and Germain, R. (2001) The free energy landscape for β hairpin folding in explicit water. *Proc Natl Acad Sci USA 98*, 14931.

[92] Rao, F. and Caflisch, A. (2003) Replica exchange molecular dynamics simulations of reversible folding. *J Chem Phys 119*, 4035.

[93] Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys 23*, 327–341.

[94] Tozzini, V. (2005) Coarse-grained models for proteins. *Curr Opin Struct Biol 15*, 144–150.

[95] Munoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A. (1997) Folding dynamics and mechanism of β-hairpin formation. *Nature 390*, 196–198.

[96] Du, D., Zhu, Y., Huang, C. Y., and Gai, F. (2004) Understanding the key factors that control the rate of β-hairpin folding. *Proc Natl Acad Sci USA 101*, 15915.

[97] Zagrovic, B., Sorin, E. J., and Pande, V. (2001) β-hairpin folding simulations in atomistic detail using an implicit solvent model1. *J Mol Biol 313*, 151–169.

[98] Dinner, A. R., Lazaridis, T., and Karplus, M. (1999) Understanding β-hairpin formation. *Proc Natl Acad Sci USA 96*, 9068.

[99] Nguyen, P. H., Stock, G., Mittag, E., Hu, C. K., and Li, M. S. (2005) Free energy landscape and folding mechanism of a beta-hairpin in explicit water: a replica exchange molecular dynamics study. *Proteins Struct Funct Bioinf 61*, 795.

[100] Garcia, A. E. and Sanbonmatsu, K. Y. (2000) Exploring the energy landscape of a β hairpin in explicit solvent. *Proteins Struct Funct Bioinf 42*, 345–354.

[101] Bussi, G., Gervasio, F. L., Laio, A., and Parrinello, M. (2006) Free-energy landscape for β-hairpin folding from combined parallel tempering and metadynamics. *J Am Chem Soc 128*, 13435.

[102] Bonomi, M., Branduardi, D., Gervasio, F. L., and Parrinello, M. (2008) The unfolded ensemble and folding mechanism of the C-terminal GB1 β-hairpin. *J Am Chem Soc 130*, 13938–13944.

[103] Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., and GM, C. (1991) A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science 253*, 657–661.

[104] Camilloni, C., Provasi, D., Tiana, G., and Broglia, R. A. (2008) Exploring the protein G helix free-energy surface by solute tempering metadynamics. *Protein Struct Func Bioinf 71*, 1647.

[105] Lindahl, E. and Hess, B. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Mod* pp. 306–317.

[106] Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc 118*, 11225–11236.

[107] Jorgensen, W. L. (1981) Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J Am Chem Soc 103*, 335–340.

[108] Evans, D. J. and Holian, B. L. (1985) The Nose–Hoover thermostat. *J Chem Phys 83*, 4069.

[109] Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J Chem Phys 98*, 10089.

[110] Yoda, T., Sugita, Y., and Okamoto, Y. (2004) Comparisons of force fields for proteins by generalized-ensemble simulations. *Chem Phys Lett 386*, 460–467.

[111] Baynes, B. M. and Trout, B. L. (2003) Proteins in mixed solvents: a molecular-level perspective. *J Phys Chem B 107*, 14058–14067.

[112] Beck, D. A. C., Alonso, D. O. V., and Daggett, V. (2003) A microscopic view of peptide and protein solvation. *Biophys Chem 100*, 221–237.

[113] Shukla, D., Shinde, C., and Trout, B. L. (2009) Molecular computations of preferential interaction coefficients of proteins. *Phys Chem B 113*, 12546–12554.

[114] Vagenende, V., Yap, M. G. S., and Trout, B. L. (2009) Molecular anatomy of preferential interaction coefficients by elucidating protein solvation in mixed solvents:

methodology and application for lysozyme in aqueous glycerol. *Phys Chem B 113*, 11743–11753.

[115] Kirkwood, J. G. and Goldberg, R. J. (1950) Light scattering arising from composition fluctuations in multi-component system. *J Chem Phys 18*, 54–57.

[116] Schellman, J. A. (1978) Solvent denaturation. *Biopolymers 17*, 1305–1322.

[117] Espinosa, J. F. and Gellman, S. H. (2000) A designed-hairpin containing a natural hydrophobic cluster. *Angew Chem Int Ed 39*, 2330–2333.

[118] Dyer, R. B., Maness, S. J., Peterson, E. S., Franzen, S., Fesinmeyer, R. M., and Andersen, N. H. (2004) The mechanism of β-hairpin formation. *Biochemistry 43*, 11560–11566.

[119] Bolhuis, P. G. (2003) Transition-path sampling of b-hairpin folding. *Proc Natl Acad Sci USA 100*, 12129.

[120] Auton, M. and Bolen, D. W. (2005) Predicting the energetics of osmolyte-induced protein folding-unfolding. *Proc Natl Acad Sci USA 102*, 15065–15068.

[121] Street, T. O., Bolen, D. W., and Rose, G. D. (2006) A molecular mechanism for osmolyte-induced protein stability. *Proc Natl Acad Sci USA 103*, 13997.

[122] Duan, Y. and Kollman, P. A. (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science 282*, 740.

[123] Zagrovic, B., Snow, C. D., Shirts, M. R., and Pande, V. S. (2002) Simulation of folding of a small α-helical protein in atomistic detail using worldwide-distributed computing. *J Mol Biol 323*, 927–937.

[124] Lei, H. and Duan, Y. (2007) Two-stage folding of HP-35 from ab initio simulations. *J Mol Biol 370*, 196–206.

[125] Mori, G. M. S. D., Micheletti, C., and Colombo, G. (2004) All-atom folding simulations of the villin headpiece from stochastically selected coarse-grained structures. *J Phys Chem B 108*, 12267–12270.

[126] Piana, S., Laio, A., Marinelli, F., Troys, M. V., Bourry, D., and Ampe, C. (2008) Predicting the Effect of a Point Mutation on a Protein Fold: The Villin and Advillin Headpieces and Their Pro62Ala Mutants. *J Mol Biol 375*, 460–470.

[127] Piana, S. and Laio, A. (2007) A bias-exchange approach to protein folding. *J Phys Chem B 111*, 4553–4559.

[128] Piana-Agostinetti, S., Lindorff-Larsen, K., Maragakis, P., Eastwood, M. P., Dror, R. O., and Shaw, D. E. (2009) Improving molecular mechanics force fields by comparison of microsecond simulations with NMR experiments. *Biophys J 96*, 406.

[129] Best, R. B., Buchete, N. V., and Hummer, G. (2008) Are current molecular dynamics force fields too helical? *Biophys J 95*, L07———L09.

[130] Vermeulen, W., Vanhaesebrouck, P., Troys, M. V., Verschueren, M., Fant, F., Goethals, M., Ampe, C., Martins, J. C., and Borremans, F. A. M. (2004) Solution structures of the C-terminal headpiece subdomains of human villin and advillin, evaluation of headpiece F-actin-binding requirements. *Protein Sci 13*, 1276–1287.

[131] Hess, B., Kutzner, C., Spoel, D. V. D., and Lindahl, E. (2008) Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput 4*, 435–447.

[132] Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem 18*, 1463–1472.

[133] Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. V., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys 81*, 3684.

[134] Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J Chem Phys 126*, 14101.

[135] Reiner, a., Henklein, P., and Kiefhaber, T. (2010) An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc Natl Acad Sci USA 107*, 4955–4960.

[136] Tang, Y., Rigotti, D. J., Fairman, R., and Raleigh, D. P. (2004) Peptide models provide evidence for significant structure in the denatured state of a rapidly folding protein: The villin headpiece subdomain. *Biochemistry 43*, 3264–3272.

[137] Kubelka, J., Eaton, W. A., and Hofrichter, J. (2003) Experimental tests of villin subdomain folding simulations, J. *Mol Biol 329*, 625–630.

[138] Kubelka, J., Henry, E. R., Cellmer, T., Hofrichter, J., and Eaton, W. A. (2008) Chemical, physical, and theoretical kinetics of an ultrafast folding protein. *Proc Natl Acad Sci USA 105*.

[139] Havlin, R. H. and Tycko, R. (2005) Probing site-specific conformational distributions in protein folding with solid-state NMR. *Proc Natl Acad Sci USA 102*, 3284–3289.

[140] Tang, Y., Grey, M. J., McKnight, J., Palmer, A. G., and Raleigh, D. P. (2006) Multistate folding of the villin headpiece domain. *J Mol Biol 355*, 1066–1077.

[141] Wickstrom, L., Okur, A., Song, K., Hornak, V., Raleigh, D. P., and Simmerling, C. L. (2006) The unfolded state of the villin headpiece helical subdomain: Computational studies of the role of locally stabilized structure, J. *J Mol Biol 360*, 1094–1107.

[142] Neumann, M. (1985) The dielectric constant of water. Computer simulations with the MCY potential. *J Chem Phys 82*, 5663.

[143] van der Spoel, D., Maaren, P. J. V., and Berendsen, H. J. C. (1998) A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *J Chem Phys 108*, 10220.

[144] Shikata, T. and Itatani, S. (2002) Dielectric Relaxation of Aqueous Trimethylamineoxide Solutions. *J Sol Chem 31*.

Acknowledgements

Prima di tutto vorrei ringraziare i miei genitori che, non senza perplessità, mi hanno sempre sostenuto lungo questa (difficile) strada, quella “di chi fa ricerca”, non sempre ricca di successi e soddisfazioni. In secondo luogo ringrazio il mio gruppo. Il mio tutor Prof. Maurizio Sironi e il Dott. Stefano Pieraccini che con me hanno condiviso i piccoli intoppi e gli importanti successi di questi tre anni di ricerca, il Dott. Stefano Rendine, che ha avuto l’arduo compito di sostituire “la coinquilina” e nonostante tutto è riuscito a rendere piacevoli tante giornate di lavoro. E naturalmente “la coinquilina” stessa che rimane presente, anche se da lontano, con la sua immane saggezza. Tutti i tirocinanti e tesisti che hanno frequentato il Molecular Modeling Group in questi tre anni, ed in particolar modo Riccardo, che ho avuto la fortuna di seguire durante la tesi triennale, e che ho la fortuna ora di vedere sempre più consapevole e in gamba nel suo lavoro. Ringrazio i miei “fratelli” dottorandi Alberto, Ivano, Simone, Elisa, Fausto, Alessandra che hanno condiviso questa esperienza con me e naturalmente gli altri amici delle pause pranzo, Cecilia, Luca, Edoardo, Silvia e tutti gli altri. Ringrazio Agostina, per le innumerevoli cene indimenticabili, per le pizze e i tiramisù, per le mille verità sotto effetto del fragolino. Ringrazio i miei amici Daniele, Matteo, Salvo, Egidio, Vincenzo e Michele per innumerevoli motivi che sarebbe troppo lungo elencare. Ringrazio il gruppo di biofisica computazionale del CNIO di Madrid: Ilaria, Ludovico, Marco, Neva, Jarek e Francesco, perchè è raro trovare persone tanto in gamba e tanto intelligenti tutte nello stesso gruppo. Ringrazio anche tutti gli altri ragazzi del CNIO che ho avuto modo di conoscere nei mesi trascorsi a Madrid, tra cui Paolo, Enza, Michele, Francesco e gli altri italiani. Me gustaría dar las gracias a todos mis compañeros españoles, que hicieron este

2010 inolvidable y que dejaron una marca permanente en mi alma. Me gustaría dar las gracias a Sara, por su desventuras, sus citas locas, su italiano extraño mezclado con palabras en español, y la hermosa ciudad de Corbolacchione. Eva y Chiara, por la presencia constante en Madrid, para las damas hermosas de c/ fuencarral y para las noches de Cluedo (y, naturalmente, para el "trentathlon"). No puedo dejar de agradecer a Tommy, que permaneció sólo tres meses, pero siempre estábamos hablando de él (y para la fiesta de cumpleaños de Nicolás... y no voy a decir nada más). Me gustaría dar las gracias a todos los amigos en Madrid, Sergio, Elena, Marco, Griselda, Nicolás y, naturalmente, mi querida vecina Mercedes.

... y naturalmente, me gustaría dar las gracias a "la señora", que dio tanto en tan poco tiempo (las fiestas por la copa del mundo, incluso) y que sigo echando de meno.