Scuola Normale Superiore di Pisa

Classe di Scienze

Large scale conformational plasticity in biological macromolecules

Tesi di perfezionamento
in chimica

Candidata:

Anna Berteotti

Relatore:

Prof. Dr. Michele Parrinello

Anno Accademico 2010/2011

Contents

Chapter 1

Introduction

Molecular flexibility is an intrinsic property of biomolecules[1]. It is crucial for such basic functions as enzyme catalysis, regulation of protein activity, transport of metabolites, etc.. Higly flexible proteins have been either identified or implicated in various diseases. Understanding the fundamental nature of biomolecular flexibility is important not only for relating structural information to biological function, but also for the development of novel therapeutics. However, studying biomolecular flexibility represents a very difficult task both experimentally and theoretically. Up to now, all information we have about the macroscopic structure of proteins come from X-ray, neutron scattering and NMR studies. However, these techniques provide only a static picture of biological molecules and a comprehension of how proteins work requires an understanding of the connection between three-dimensional structure and dynamics, which is much more difficult to probe experimentally. Although significant results have been obtained by NMR relaxation experiments as well as by single-molecule fluorescence resonance energy transfer (FRET) measurements, these experiments provide ensemble- or time-averaged data and translating them into an atomic detailed view is generally not trivial. In this scenario, atomistic simulations can in principle bring an invaluable insights. Molecular Dynamics (MD) is a precious instrument to understand the mechanism underlying complex processes, providing the linkage between experimental structure with dynamics[2]. Although classical MD

simulation is nowadays a standard tool because it provides a very detailed representation of protein dynamics, it still faces significant limitations. The force field, able to represent intra-molecular and inter-molecular interactions does not quite have the required accuracy, the atom motion is described by classical dynamics, the formation and breaking of chemical bonds is not allowed, electrons are confined to remain in the ground state, and size and time scale are severely restricted. There is an increasing amount of work aimed at overcoming these limitations: more potent computers come together with the development of scalable codes, more accurate force field are being developed, bigger system are being simulated, transitions on longer time scales are being investigated. Within these limitations the most severe one is the time scale problem. This problem arises from the fact that the typical integration time of an atomistic simulation is in the fs range, because it must be smaller than the fastest degree of freedom one wants to describe. This limits current large-scale simulations where most phenomena of interest as the protein conformational plasticity or protein folding generally take milliseconds or longer times. To overcome this limitation, a huge variety of different methods have been proposed over the course of the last few years. Within these methods our group has introduced a powerful method, called metadynamics[3]. In metadynamics, a history dependent potential is added in the space of the collective variables (CV). This forces the dynamics to explore conformations that were not previously visited and discourages the system from previously explored regions. Therefore, it allows the system to escape minima along low free energy paths and exploring other minima in the free energy landscape. The free energy surface can be obtained as the negative of the potential added during the simulation.

In this framework, the two main topics presented in this thesis, represent two interesting case of large scale motion in biology, studied by means of metadynamics simulations. The more advanced "state of the art" of metadynamics[4] has been applied in order to better understand the problem of kinase conformational plasticity and the open issue of folding/unfolding equilibrium of protein in water/urea mixed solvent. Before coming to the outline of the thesis I will give a brief introduction about the two main topics of the present dissertation.

1.1 Conformational changes in protein kinase

Protein kinases represent one of the biggest superfamily in the human genome[5]. All protein kinases catalyze the same reaction: the transfer of gamma phosphate from ATP to a peptidic substrate. Since the phosphorilation cascade plays a central role in the regulation of cell growth and differentiation, dysregulation of kinase activity can result in dramatic changes in the control of these processes.

For their relevant role in human body, protein kinases are intriguing drug targets and in the last decade much effort has been put by pharmaceutical industry in designing new potent and selective inhibitors able to interfere with their activity[6]. Such inhibitors have significant potential in the clinical treatment of major diseases including cancer, heart disease, inflammatory disorders, diabetes and neurological disorders. These inhibitors alter the function of kinase and in principle a good kinase inhibitor has to rapidly and selectively interfere with the activation state of the targeted kinase. In fact the over 1200 solved structures of protein kinase domain allow depicting a general framework of kinase activation profile[7]. Although what I'm going to describe is a general picture, based on static "images" of the proteins, it provides invaluable insights into the molecular details of kinase function and regulation, moreover it is fundamental for any kinase drug discovery project[8]. We can think of protein kinases as molecular switches that can adopt at least two conformations[9, 10]: an "active" state that is maximally active and an "inactive" state that has minimal activity. Since protein kinases catalyze the same reaction, upon activation, they adopt catalytically active conformations that are structurally very similar since they have to perform similar tasks. The protein kinase fold, which is extremely conserved among kinases, is composed by two subdomains or lobe connected by a semi-flexible hinge region (see Fig. 1.1). The N-terminal lobe, or N-lobe, is mostly composed by β-sheets. However, it includes one α-helix, called αC-helix. The C-terminal lobe, or C-lobe is mostly α-helical and contains the activation segment. The first three residues of the activation segment form a highly conserved Asp-Phe-Gly triad known as the "DFG" motif. The activation segment contains hydroxyl-bearing amino acids residues that are phosphorilated to activate the

enzyme for catalysis. Between the two lobes, there is the ATP-containing site where the substrate binds. The ATP-binding site sits beneath a highly conserved loop connecting strands two β-sheets. This loop, called G-loop, contains a conserved glycine-rich sequence that allow the loop to approach the phosphates of ATP very closely and to coordinate them via backbone interactions.

In the active state, the conserved Glu in the αC-helix forms a salt bridge with a lysine residue of the N-lobe. This salt bridge is essential for catalysis and is highly conserved among all protein kinases (see Fig.1.2A). Also the Asp, inside the DFG motif, points into the ATP binding site and the activation segment is in a open and extended conformation.

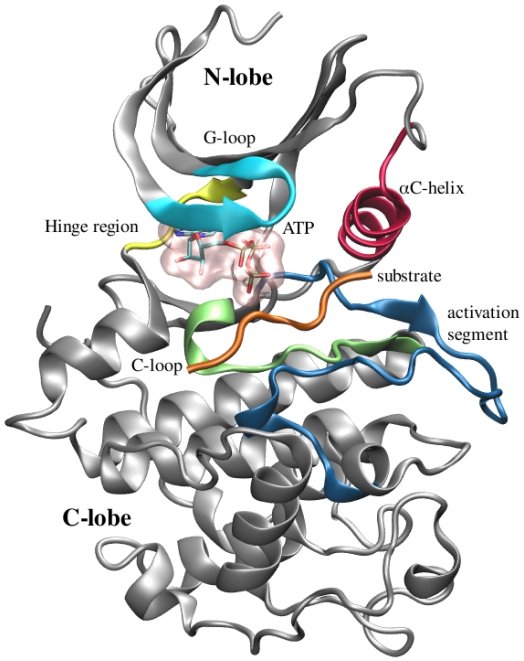

Figure 1.1: Active state of Cyclin dependent kinase 2 with ATP as licorice and the peptide substrate in orange. The classical bilobal structure of kinase can be seen: the G-loop (cyan); the αC-helix (violet); the hinge region between the N- and C-lobes of the protein (yellow); the C-loop (green); the activation segment (blue).

On the other hand, the inactive state can be different reflecting the fact that the

catalytic activity can be blocked in different ways. For what concern the inactive states, there is a large variety in the solved structures. They can be divided into two main groups (see Fig. 1.2), corresponding to the different classes of inhibitors that they can bind:

- in the first type, that we can call αC-helix Glu-Out, the conserved salt bridge that anchors the αC-helix in active structures is broken and the αC-helix moves away from the ATP-binding site, creating an additional pocket such as in a tyrosine kinase, the epidermal growth factor, and this pocket can be occupied by an inhibitor such as Lapatinib[11];
- in the second type, the Asp sidechain in the conserved DFG motif is rotated out of the ATP binding site, creating a new pocket. We can referee to this inactivation type as DFG-Out. This pocket can be occupied by selective inhibitors, including the marketed drugs Imatinib[12] and Sorafenib[13].

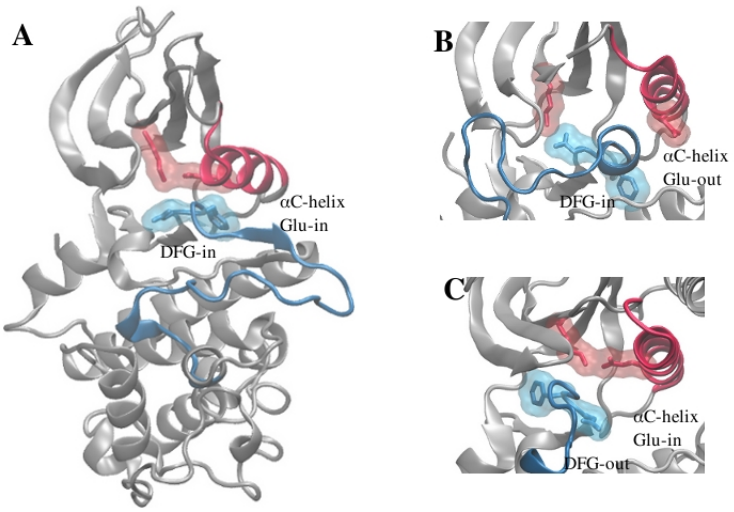

Figure 1.2: A: Active state of CDK2 kinase with DFG-in conformation and αC-helix Glu-in, B: inactive state of CDK2 with DFG-in and αC-helix Glu-out conformation, C: P38α Map kinase with DFG-out and αC-helix Glu-in conformation

In both groups, the highly flexible activation segment can be in a substantial different position compared to its place in the active state. Up to now, only few kinases[14, 15] have

been seen to adopt both the DFG-out and the αC-helix Glu-out inactive conformations. This underlines the importance of considering the X-ray structures as snapshot which depicts a kinase at a particular point of a conformational equilibrium rather than be taken as a global image of the active or the inactive state. Moreover it clearly emerges the limitation of a two-state representation of kinase activation, while it is more realistic a view in which the protein can be in conformational equilibrium between several states. This equilibrium can be shifted to one or another state in particular environment conditions.
In this thesis I will present our results about the possibility to simulate the process of inactivation for both the inactive classes mentioned above. The transition from the active, αC-helix Glu-in, to the inactive conformations of the first type, αC-helix Glu-out, has been studied for Cyclin Dependent Kinase 5[16, 17] (CDK5). The second type of inactivation process has been studied for Mitogen-activated protein kinase p38 α[18] (p38). In this case we have simulated the transition from DFG-in to DFG-out state, we have also studied the role of an inhibitor on the conformational transition between the two states. In both cases we have also computed the free energy surface associated to this event. Moreover we have found out several intermediate states going from active to inactive states that can be used in further drug design project in order to design new selective inhibitors.

1.2 Urea-Induced Protein Denaturation

Herafter I introduce the second main topic of the present dissertation: the folding-unfolding mechanism in pure water and in 8M urea solution.
A protein remains folded as a result of a delicate balance of forces between the protein's interactions with itself and its interactions with its environment. Under any given conditions, protein molecules in aqueous buffer are in equilibrium between unfolded and folded states, U(nfolded)$\rightleftharpoons$N(ative)[19]. This balance can be perturbed in different ways, by heat, cold, pressure, acid, alkali, or organic molecules. Small uncharged molecules, called osmolytes, are used by the cell to avoid an osmotic catastrophe. In fact the cell volume is maintained in osmotic equilibrium via carefully controlled changes in the intracellular concentrations

of these osmolytes. They may be classified as either denaturing or protecting. In the equilibrium folding reaction, U⇌N the denaturing osmolyte, as for example, urea stabilizes U relative to N; protecting osmolytes, as Trimethylamine N-oxide (TMAO), do the opposite, stabilizing N relative to U.

Urea, a small hydrophilic molecule, present in all taxa, is a widely used protein denaturant in vitro unfolding/refolding experiments. However, despite its everyday use in studying protein folding and stability, the molecular mechanism by which urea acts as denaturant is still not well understood. In particular, the question of which kind of interactions embody the main driving force inducing denaturation has been intensively studied. Two main different mechanisms have been proposed:

1. an "indirect mechanism" in which urea is presumed to disrupt the structure of water, thus making hydrophobic groups more readily solvated[20, 21, 22];

2. a "direct mechanism" in which urea can interact either directly with the protein backbone, via hydrogen bonds and other electrostatic interactions, or directly with the amino acids through more favorable van der Waals attractions, or with both, thus causing the protein to open, and then denature[23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34].

Within this debate, MD simulation could be of great importance in order to elucidate the mechanism of urea denaturation. However, recent MD simulations have led to qualitatively different conclusions. Many MD simulation have sought to understand the mechanism by focusing on the pathway of denaturation rather than examining the effect of urea on the folding/unfolding equilibrium. In addition, although urea shifts the U⇌N equilibrium toward the unfolded state, the accessible computer time does not allow a simulation of the entire unfolding event. Thus only partial unfolding events, even at elevated temperature, have been reported in many studies. Lately time scale of the simulations has been increased to a few microseconds, and in lysozyme a two step process has been found[35]. While highly illuminating, these calculations did not report any free energy calculations. Efforts at improving sampling has also been recently made by Canchi et al.[36]

who have used replica exchange to study the folding equilibrium of the Trp cage protein in the presence of urea finding, in agreement with experiments, a linear dependence of unfolding free energy on urea concentrations.
In this scenario, I will report in this thesis our computational study on folding/unfolding equilibrium of the 16-residue C-terminal fragment of protein GB1 (β-hairpin) in 8M urea compared to pure water solution. The use of advanced sampling techinques has allowed us to analyze the mechanism of urea and to study the nature of the unfolded state in both solutions.

1.3 Outline of the thesis

Beside this brief introduction, this dissertation is organized in four main chapters.
Chapter 2 is dedicated entirely to the basic theory of the simulation techniques used in this thesis. Firstly I will briefly describe the MD technique applied to biological system. Then I will focuse on the theory of metadynamics. The details of the most used form, the so-called direct formalism, as well as its latest variant, well tempered metadynamics, are explained. The pros and cons of metadynamics are described with a particular focus on the collective variables. Then I introduce the theory of Path Collective Variables (PCV) that are able to describe complex reactions with a limited set of CVs. Lastly I show how is it possible to combine metadynamics with parallel tempering to mitigate the problem of including all the slow modes of the system in the set of CVs.
Chapter 3 illustrates the application of metadynamics within the PCV to the problem of large scale motion in kinases. This chapter presents the atomistic dynamics of the "open-to-close" movement of the CDK5. I briefly introduce the open to closed problem in kinases, then I will illustrate the mechanism of that transition for the protein under investigation. I also show the estimated free-energy profile associated with the global motion and the presence of an intermediate structure in between the ending ones, which could be exploited for drug-design purposes.
Chapter 4 is concerned with the second type of kinase inactivation process, studied in

p38 kinase. This mechanism involves the rotation of about 10 Å of the DFG motif. I will show the mechanism of this transition for the protein alone and in the presence of an inhibitor. I will describe how this inhibitor interferes with the inactivation mechanism.

Chapter 5 deals with the problem of protein unfolding in different environments such as simple water and urea-water mixture. In this chapter I will present the mechanism by which urea is able to interact with the β-hairpin protein and to stabilize the unfolded state.

Chapter 2

Theoretical Background

2.1 Molecular Dynamics

Since the first report of MD simulation of protein some 30 years ago[37], MD has become an established tool in the study of biomolecules. It is used to understand the mechanism underlying complex processes, both for interpreting experimental data and for making novel predictions.

MD consists of observing the evolution of a system of N particles, under the action of some forces arising from their interactions and under certain condition of temperature, pressure and volume. Given the positions and the velocities at an initial instant t_0, the position of each particles $\boldsymbol{R}_j$ is updated after a small time step by integrating the Newton's equation of motion:

$$M_j \ddot{\boldsymbol{R}}_j = -\nabla_j U(\boldsymbol{R}) \; \forall j = 1, ..., N \tag{2.1}$$

where $\boldsymbol{R} = (\boldsymbol{R}_1, ..., \boldsymbol{R}_N)$ is the configuration of the system, $U(\boldsymbol{R})$ is the potential energy function and M_j the mass of j particle.

If the simulation is prolongated for a sufficiently long time, the system visits several microscopic configurations that can be used to measure different observables and also link the simulation with experiment. Let us consider a simulation performed at constant number of particles N, temperature T and volume V (canonical or NVT ensemble). An observ-

able can be defined as the ensemble average of a function $s = s(\boldsymbol{R})$ of the microscopic coordinates:

$$\langle s \rangle = \int ds\, s\, P(s) \tag{2.2}$$

where $P(s)$ is the probability distribution of s in the canonical ensemble. This can be written as:

$$P(s) = \langle \delta(s - s(\boldsymbol{R})) \rangle = \frac{\int d\boldsymbol{R}\, \delta(s - s(\boldsymbol{R})) e^{-\beta U(\boldsymbol{R})}}{Z} \tag{2.3}$$

being $\beta = 1/(k_B T)$, k_B the Boltzmann constant and Z the partition function:

$$Z = \int d\boldsymbol{R}^{-\beta U(\boldsymbol{R})} \tag{2.4}$$

The probability distribution $P(s)$ is related to the Helmholtz free energy by the relation:

$$F(s) = -\frac{1}{\beta} \ln P(s) + C \tag{2.5}$$

where C is an irrelevant additive constant. Since the configurations produced in a MD simulation in NVT conditions are distributed according to the Boltzmann weight $e^{-\beta U(\boldsymbol{R})}$, one can accumulate in the histogram $N(s)$ the variable s, during a long MD run in which all the energetically relevant configurations have been sampled (ergodic hypothesis, see section 2.2); one can also estimate the probability distribution $P(s)$ using the normalized histogram $N(s)$; or calculate the ensemble average using Eq. 2.2. Alternatively, if one is not interested in $P(s)$, the ensemble average s can also be calculated directly from the time average:

$$\langle s \rangle = \frac{1}{N} \sum_{i=1}^{N} s(\boldsymbol{R}(t_i)) \tag{2.6}$$

where N is the number of configurations collected during a long MD run, $\boldsymbol{R}(t_i)$ the configuration at time t_i.

Defining the potential energy

Depending on the way how forces are computed, it is possible to classify the MD methods as atomistic or *ab-initio*, where instead of describing the interaction potentials by analytical functions the electronic structure is treated explicitly. A great advantage of *ab-initio* MD

is that it allows bond building/breaking reactions and excited states to be described. In force field based MD the interactions between atoms are described by suitable combinations of analytical functions, which are fitted either on *ab-initio* calculations or parametrized by experimental data. The use of analytical functions allows a fast evaluation of the interatomic interactions, and therefore large systems can be sampled in the nanosecond regime. However, the choice of the parameters is essential to set up a model which is able to reproduce realistically the system of interest. In the following we shall focus on the simulation of biophysical processes involving proteins, such as protein folding or protein-protein interactions, at physiological conditions. For this purpose, a variety of possible choices can be made for the force field. Here we limit our attention to force fields that describe the protein and the solvent degrees of freedom at an atomistic detail.

The choice of the time step

In MD the integration of Newton's equation of motion is based on the discretization of a second order differential equation. Whatever algorithm one might use to integrate the equations of motion, the maximum time step to be used safely is determined by the frequency of the fastest motion in the system. For molecular systems these are the stretching vibrations involving the hydrogen atoms, leading to time steps of the order of femtoseconds. For proteins this time scale is orders of magnitude smaller compared to the important motions, such as those involved in functionally important conformational changes or in folding, which are in the range of microseconds to seconds or even longer. A small time step ensures a good energy conservation and a good accuracy for the sampling of the statistical ensemble, but on the other hand this leads to a huge number of steps needed to sample the phase space.

The problem of rare events

The possibility to extract the probability distribution from a MD run is reliable only if the MD run is long enough for the system to visit all the energetically relevant configurations or, in other words, if the system is ergodic in the time scale of the simulation. In

practical situations this rarely happens. One reason is because the relevant configurations are often separated by high free-energy barriers. In such cases one should wait for thermal fluctuations to activate the reaction, i.e. to drive the system above the energy barrier. A second possibility is that the system diffuses extremely slowly in the intermediate regions. For all these reasons, such processes are often referred to as "rare events".

To accumulate sufficient statistics on these events, a straightforward MD simulation is almost useless when carried out for a finite amount of time. Since in biophysical problems we used a time step in the order of femtoseconds and relevant processes like folding/unfolding, conformational movement are in the range of microseconds to seconds or even longer, it emerges clearly that MD is afflicted by a problem of time scale. To deal with this problem, a huge variety of different methods has been proposed over the course of the last few years. Here we focus on a method, called metadynamics[3], that belongs to a more general group of methods, in which the system experiences an additional potential acting on a selected number of degrees of freedom, often referred as collective variables. $\boldsymbol{S}$ is a set of d functions of the microscopic coordinate $\boldsymbol{R}$ of the system. The additional time-dependent potential $V(S(\boldsymbol{R}), t)$, modifies Newton's equation of motions as:

$$M\ddot{\boldsymbol{R}} = -\frac{\partial U(\boldsymbol{R})}{\partial \boldsymbol{R}} - \frac{\partial V(S(\boldsymbol{R}), t)}{\partial \boldsymbol{R}} \tag{2.7}$$

In metadynamics a history-dependent bias potential that depends on the microscopic variables through its dependence on $\boldsymbol{S}$ is used. The bias potential allows the system to be pushed out of its initial metastable state and accelerates the exploration of other metastable states. We focus here on metadynamics, in its standard form or in the recently introduced well-tempered flavor[4]. In particular, we consider the direct version of metadynamics, where the bias is acting directly on the microscopic coordinates.

2.2 Metadynamics

In standard metadynamics the bias potential is built during the simulation as a sum of Gaussian functions centered on the previously visited configurations in the CVs space. Let

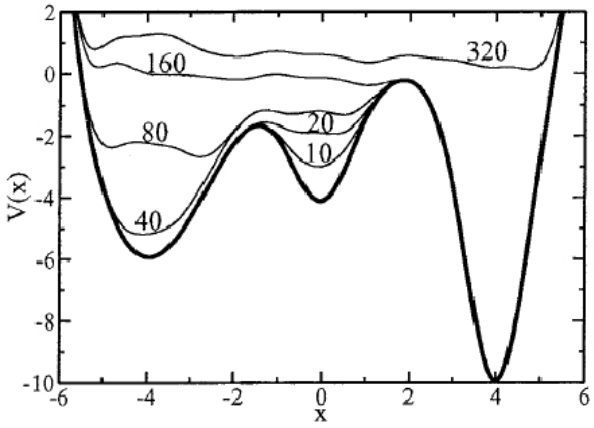

Figure 2.1: Time evolution of the sum of a one-dimensional model potential V (x) and the accumulating Gaussian terms of Eq. 2.9. The dynamic evolution (thin lines) is labeled by the number of Gaussian added. The starting potential (thick line) has three minima and the dynamics is initiated in the second minimum. The figure is taken from Ref.[3].

$\boldsymbol{S}$ be a set of d functions of the microscopic coordinate $\boldsymbol{R}$ of the systems:

$$\boldsymbol{S}(\boldsymbol{R}) = (\boldsymbol{S}_1(\boldsymbol{R}), ..., \boldsymbol{S}_d(\boldsymbol{R})). \tag{2.8}$$

In *so-called* continuous and direct metadynamics an external history dependent bias potential $V_G = V_G(\boldsymbol{S}, t)$ function of these CVs is added to the Hamiltonian of the system and acts directly on the microscopic coordinates. This potential can be written as a sum of Gaussian deposited along the trajectories of the CVs:

$$V_G(\boldsymbol{S}, t) = \int_0^t dt^{'} \omega \exp\left(-\sum_{i=1}^{d} \frac{(S_i(\boldsymbol{R}) - S_i(\boldsymbol{R}(t^{'})))^2}{2\sigma_i^2}\right) \tag{2.9}$$

where σ_i is the width of the Gaussian for the ith CV and ω is the rate at which the bias grows. The energy rate is constant and is usually expressed in terms of a Gaussian height W and a deposition stride τ_G:

$$W = \frac{\omega}{\tau_G} \tag{2.10}$$

To understand the effect of V_G on the evolution of the system, let us consider the simple case of the one-dimensional potential described in Fig. 2.1.

In this example we shall measure the time t by counting the number of Gaussians deposited. The system is initialized in one local minimum, say the second one. As time goes by, Gaussians are deposited at present position. The underlying bias potential grows until the system is pushed away to a new local minimum. This happens around $t = 20$. The natural and more convenient escape route is passing above the lowest barrier to fall into the left basin. Here the Gaussians accumulation starts again. The system is trapped in this local minimum until the underlying free-energy basin is completely filled ($t \simeq 120$). Starting from $t = 160$ the system can easily access also the third minimum region on the right. Finally, when also this basin is compensated by the bias potential ($t = 320$), the system can move in a random walk on the flat free energy surface (FES). Metadynamics has been historically used in two different and complementary manners. It has been used to escape free-energy minima, i.e. to find a reasonable saddle point out of a local minimum. In this case, metadynamics should be stopped as soon as the system exits from the minimum and starts exploring a new region of space. In other applications, it has been used to exhaustively explore the CV space and reconstruct the FES. Here I focus more on this latter application.

The example shows clearly the multiple benefits of metadynamics: it accelerates the sampling of rare events by pushing the system away from local free-energy minima; it allows the exploration of new reaction pathway as the system tends to escape these minima passing through the lowest free-energy saddle point; no a priori knowledge of the landscape is required. At variance with umbrella sampling, metadynamics inherently explores first the low free-energy regions; the sum of all the Gaussians deposited in the long time provides an estimate of the underlying FES. As a matter of fact, it has been shown analytically that V_G is an unbiased estimator of the free energy $F(S)$:

$$\lim_{t\to\infty} V_G(\boldsymbol{S}, t) = -F(\boldsymbol{S}) + C \tag{2.11}$$

where C is an irrelevant additive constant.

Two main drawbacks, however, are connected to metadynamics:

1. V_G does not converge to the free energy, but oscillates around it. This fact has

two consequences: a) the bias potential overfills the underlying FES and pushes the system to visit high-energy regions of the phase space; b) it is difficult to decide when to stop a simulation. In this regard, if metadynamics is used to find the closest saddle point to a given local minimum, it should be stopped as soon as the system exits from that minimum. Otherwise, if one is interested in reconstructing a FES, metadynamics should be stopped when the motion of the CVs becomes diffusive in the region of interest;

2. the CVs must be chosen a priori and this choice is often far from trivial. In section 2.4 I describe the criteria that define a good set of CVs, the most important being that all the slow modes of the system must be included. Unfortunately, while it is quite easy to realize a posteriori that a set of CVs is bad, it is hard to identify the proper set of CVs prior to the simulation.

Before discussing the choice of the CVs I describe how well-tempered metadynamics, can help in solving the first drawback of metadynamics.

2.3 Well-tempered metadynamics

In well-tempered metadynamics[4] the initial deposition rate ω_0 decreases with the bias accumulated over time. This is achieved by rescaling the Gaussian height W according to:

$$W = \omega_0 \tau_G e^{-\frac{V_G(\boldsymbol{S},t)}{k_B \Delta T}} \tag{2.12}$$

where τ_G is the Gaussian deposition stride, ΔT a temperature and $V_G(\boldsymbol{S},t)$ is the bias potential accumulated in $\boldsymbol{S}$ over time t. At variance with standard metadynamics, the bias potential does not fully compensate the FES, but it converges to:

$$V_G(\boldsymbol{S}, t \to \infty) = -\frac{\Delta T}{\Delta T + T} F(\boldsymbol{S}) \tag{2.13}$$

where T is the temperature of the system. In other words, at convergence the CVs are sampled at a (fictitious) higher temperature $T + \Delta T$:

$$P(\boldsymbol{S}, t \to \infty) \propto e^{-\frac{F(\boldsymbol{S})}{k_B (T + \Delta T)}} \tag{2.14}$$

Well-tempered metadynamics solves two of the major problems of standard metadynamics. The first one is the lack of convergence. In standard metadynamics the bias is continuously added to the system, also after compensating the underlying FES. This causes the metadynamics estimate of the free energy to oscillate around the correct value. In well-tempered metadynamics the amount of bias added decreases in time, its variation going to zero as $1/t$ while V_G converges to a fraction of the FES (Eq. 2.13). The second problem, strictly related to the first one, is overfilling. In well-tempered metadynamics, the exploration can be restricted to low free-energy regions by properly tuning ΔT. In particular, standard MD is recovered for $\Delta T \to 0$, traditional metadynamics for $\Delta T \to \infty$. This possibility of regulating the extent of exploration is particularly useful for saving computational time in case a large number of CVs is used.

2.4 The collective variables

Metadynamics permits an efficient calculation of $F(\boldsymbol{S})$ once an appropriate set of $\boldsymbol{S}$ has been chosen. A crucial question is how to choose the $\boldsymbol{S}$. A collective variable is a function of the microscopic coordinates of the system on which the bias potential of metadynamics acts[38]. Ideally the CVs should satisfy three properties:

- They should distinguish among the initial, the final states and all the relevant intermediates: in fact if they cannot discriminate between the configurations of the reactants, of the transition state, of the products, and of the relevant intermediates, they are not appropriate for studying a particular reaction.

- They should include all the slow events that are relevant to the process under investigation: this prerequisite is essential. We define as “slow” those variables that present, at a given temperature, free-energy barriers so high that cannot be easily overcome in the time scale of our simulation. We call fast” those variables that can be satisfactorily sampled. If any of the slow variables is not added to the CVs list, the bias potential may not converge to the FES in a reasonable simulation time. Let

us consider an example and examine the two-dimensional Z-shaped potential of Fig. 2.2.

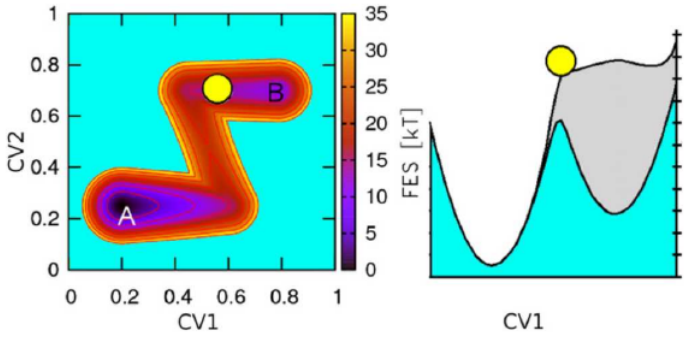

Figure 2.2: The effect of neglecting a relevant degree of freedom. Left side: 2D Z-shaped potential. Right side: the trajectory of a metadynamics simulation generated using only cv1 as collective variable. Transitions from A to B are not properly described by cv1, causing strong hysteresis in the reconstructed free energy. The figure is taken from Ref.[38].

The transition between the two metastable states A and B presents high free-energy barriers in both cv1 and cv2. We shall start from basin B and bias only cv1. When metadynamics fills the basin, the system remains stuck there, as it faces a high barrier also in the hidden variable cv2. As a result, basin B is overfilled and the barrier for going to A is overestimated. Eventually, cv2 makes the transition and the system reaches state A. At this point, a similar phenomenon occurs again. The overall result is that the situation in which the free energy grows evenly and the system diffuses from A to B is never reached. A similar behavior is often observed in real cases and is a strong indication that an important CV is missing. The usual warning is a characteristic hysteretic behavior in which, once the system has moved from one basin to the other, it has difficulty in coming back to the initial metastable state. A deeply observation of the system under investigation will show which are the slow variables not included in the CVs. This will allow not only a proper reconstruction of the $F(\boldsymbol{S})$ but, more important, a fully understanding of the underlying physics will be gained.

- They should be limited in number, otherwise it will take a very long time to fill the free energy surface. Of course, the number of CVs should be kept small. Although with well-tempered metadynamics this problem should be alleviated, the use of many CVs implies that a high-dimensional space has to be explored. This may take a considerable amount of computational time, with the additional risk that the resulting FES may not be as readily comprehensible as a low-dimensional one.

In order to alleviate these limitations of metadynamics, two methods that can help in a multitude of situations, have been developed in the last recent years:

1. the Path Collective Variables (PCV) are a set of flexible descriptor that represent a configuration in terms of a progress along a reference path connecting two state A and B, and a distance from it. These collective variables are very useful in case of multidimensional problem;

2. the combination of Parallel Tempering and Metadynamics (PTMETA) has been developed for those case in which some slow modes of the system have not been included in the set of CVs.

2.5 Path collective variables

Consider a transition between the stable or metastable state A and B. Assume that this transition can be described by a set of collective variables $\boldsymbol{S}(\boldsymbol{R})$ which are in general a multidimensional vectorial functions of the microscopic variables $\boldsymbol{R}$. Contrary to metadynamics, the dimension of the vector $\boldsymbol{S}$ is not restricted to a small number but can be arbitrarily large. If the choice of $\boldsymbol{S}$ is appropriate, we would expect the reactive trajectories to be bundled in a narrow tube around a path, which we can write in parametric form as $\boldsymbol{S}$ for $0 \leq t \leq 1$ with $\boldsymbol{S}(0) = \boldsymbol{S}(A)$ and $\boldsymbol{S}(1) = \boldsymbol{S}(B)$. We are interested in the minimum free energy path connecting state A and B on the free-energy $F(\boldsymbol{S})$ defined as in Eq.2.5:

$$F(\boldsymbol{S}) = -\frac{1}{\beta} \ln\langle\delta(\boldsymbol{S} - \boldsymbol{S}(\boldsymbol{R}))\rangle \tag{2.15}$$

where the average is taken over the Boltzmann distribution. In order to trace this path, Branduardi et al.[39] have introduced two variables:

$$s(\boldsymbol{R}) = \lim_{\lambda\to\infty} \frac{\int_0^1 te^{-\lambda\|\boldsymbol{S}(\boldsymbol{R})-\boldsymbol{S}(t)\|^2}dt}{\int_0^1 e^{-\lambda\|\boldsymbol{S}(\boldsymbol{R})-\boldsymbol{S}(t)\|^2}dt} \tag{2.16}$$

$$z(\boldsymbol{R}) = \lim_{\lambda\to\infty} \frac{1}{\lambda} \ln \int_0^1 e^{-\lambda\|\boldsymbol{S}(\boldsymbol{R})-\boldsymbol{S}(t)\|^2}dt \tag{2.17}$$

where $\|\ldots\|$ is the metric that defines the distance between two configurations. For any microscopic configuration $\boldsymbol{R}$, $s(\boldsymbol{R})$ and $z(\boldsymbol{R})$ measure its intercept and distance from the path $\boldsymbol{S}(t)$ respectively. In practical applications we describe the path with a discrete number of frames $\boldsymbol{S}(l), l = 1, P$ with $\boldsymbol{S}(1) = \boldsymbol{S}_A$ and $\boldsymbol{S}(P) = \boldsymbol{S}_B$. The integrals in equations 2.16 and 2.17 are approximated by the finite sums:

$$s(\boldsymbol{R}) = \frac{1}{P-1} \frac{\sum_{i=1}^{P}(l-1)e^{-\lambda\|\boldsymbol{S}(\boldsymbol{R})-\boldsymbol{S}(t)\|^2}}{\sum_{i=1}^{P} e^{-\lambda\|\boldsymbol{S}(\boldsymbol{R})-\boldsymbol{S}(t)\|^2}} \tag{2.18}$$

$$z(\boldsymbol{R}) = -\frac{1}{\lambda} \ln(\sum_{i=1}^{P} e^{-\lambda\|\boldsymbol{S}(\boldsymbol{R})-\boldsymbol{S}(t)\|^2}) \tag{2.19}$$

As shown in Fig. 2.3, the coordinate $s(\boldsymbol{R})$ measures progress along the reactive coordinate, while $z(\boldsymbol{R})$ is its distance from the path.

The set of these two coordinates can be used as a two-dimensional collective variable, and using for instance well-tempered metadynamics or any other free energy method we can reconstruct $F(s, z)$, which is the free energy surface as a function of s and z. Care must be taken that $\boldsymbol{S}(l)$ are equally spaced relative to the metric used and that there is sufficient overlap between the clusters defined by the reference frames.

A distinguishing feature of this method is its non-local character. When used together with metadynamics, PCV are able to find transition paths that are rather different from the guessed one. This is achieved by exploring the free energy dependence on $z(\boldsymbol{R})$, which is the variable that measures the distance from the reference path. In Ref.[39] the authors have shown that $F(s, z)$, with a good guess of $\boldsymbol{S}(t)$, offers precious information on the

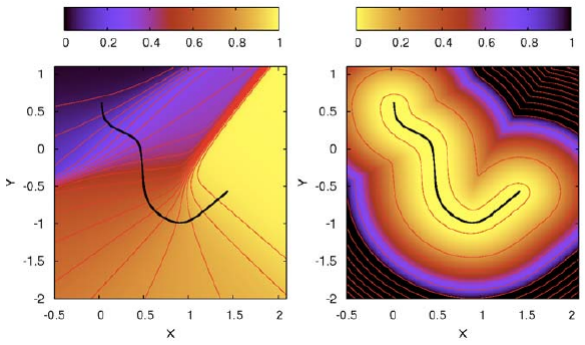

Figure 2.3: Left side: Representation of $s(\boldsymbol{R})$ and $z(\boldsymbol{R})$ isolines in a simple case potential. The $s(\boldsymbol{R})$ variable traces the closest point on the path. Right Side: the $z(\boldsymbol{R})$ variable measures the distance from the reference path.

possible reaction paths. Furthermore, an efficient procedure was proposed to improve the initially guessed path until it lies on a minimum free-energy line that connects A with B on the FES. In the following we will call this procedure optimization of the path.

From the description of PCV emerges clearly the benefit of these type of CVs:

1. While $F(s, z)$ is two-dimensional, the underlying space of the $\boldsymbol{S}$ can have very large dimensionality, thus removing the curse of dimensionality.

2. The $F(s, z)$ carries a large amount of relevant information: without any other analysis, much can be learned about the low free energy paths that join A to B by simply inspecting $F(s, z)$. We can immediately judge whether or not the initial path is good guess. In fact in the low z region, which is the closest to the path, $F(s, z)$ must be small because we expect the lowest free energy path $F(s, z = 0)$ to be a line of local minima.

3. Any other low energy path will reveal itself as a low energy valley in $F(s, z)$ and can be identified and analyzed.

4. Furthermore, by exploring the high z region of the (s, z) plane we can move away from the initial guess and discover qualitatively new transition pathways, which can be used as new guesses and similarly optimized.

The metric

The distance square $\|...\|^2$ in equations 2.16 and 2.17 can be defined in different spaces. A possible simple metric is the Root Mean Square Displacement (RMSD) between the two structures after they are optimally aligned. Different choices are also possible, as e.g. the contact map (CMAP) matrix[40]. In this case $\boldsymbol{S}(\boldsymbol{R})$ is given by the $j > i$ elements of the contact map matrix ($\boldsymbol{C}(\boldsymbol{R})$) defined as:

$$\boldsymbol{C}(\boldsymbol{R}_{i,j}) = \frac{1 - (\frac{r_{i,j}}{r_0})^p}{1 - (\frac{r_{i,j}}{r_0})^q} \tag{2.20}$$

where $r_{i,j}$ is the distance between the ith and jth couple of atoms within the protein. Proper values of of p, q, r_0 are to be set. This definition of contact map is different from the one commonly used in literature[41], which is discrete and where r_0 is intended as a sharp cutoff. Here we need to define a contact in terms of a differentiable function with continuous derivative as metadynamics requires smooth forces in the full range of distances. The distance square $\|...\|^2$ between a generic state $\boldsymbol{R}$ and a point along the path described by the CMAP $\boldsymbol{S}_C(l)$ is measured as:

$$\|\boldsymbol{S}_C(\boldsymbol{R}) - \boldsymbol{S}_C(\boldsymbol{l})\|^2 = \sum_{j>i} (\boldsymbol{C}_{i,j}(\boldsymbol{R}) - \boldsymbol{C}_{i,j}(\boldsymbol{l}))^2 \tag{2.21}$$

where the nearest neighbors are excluded from the sum.
The use of these different metrics will be seen in practical studies in Chapter 4 and 5 where the inactivation mechanism for two kinases will be described.

2.6 Parallel tempering metadynamics

In order to overcome metadynamics drawback about the difficult way of choosing the proper CVs, in 2006 it has been introduced the combination of metadynamics with parallel tempering[42]. The combined method, called PTMETA, is particularly helpful in those cases in which some slow modes of the system have not been included in the set of CVs. In parallel tempering (PT), multiple copies of the same system are simulated independently and at different temperatures. At fixed intervals, an exchange of configurations between

two adjacent replicas is attempted on the basis of a Metropolis acceptance criterion. By exchanging with higher temperatures, colder replicas are prevented from being trapped in local minima. Let us consider M independent replicas of the system at different temperatures $T_1, ..., T_M$. The configurations $\boldsymbol{R}$ of the ith replica are distributed according to the Boltzmann distribution at temperature T_i:

$$P_i(\boldsymbol{R}) = \frac{1}{Z_i} e^{-\beta_i U(\boldsymbol{R})} \tag{2.22}$$

where $\beta_i = 1/(k_B T_i)$, $U(\boldsymbol{R})$ is the potential energy function and Z_i the partition function:

$$Z_i = \int d\boldsymbol{R} e^{-\beta_i U(\boldsymbol{R})} \tag{2.23}$$

The probability distribution of the extended system of M non-interacting replicas is defined as:

$$P(\boldsymbol{R}_1, \beta_1; ...; \boldsymbol{R}_M, \beta_M) = \prod_i^M P_i(\boldsymbol{R}_i). \tag{2.24}$$

We indicate with $p(j \rightarrow k)$ the probability that a configuration $\boldsymbol{R}_j$ of the jth replica is exchanged with a configuration $\boldsymbol{R}_k$ of the kth replica. Following this notation, the probability of the inverse process is $p(k \leftarrow j)$. In order for the extended system to be at equilibrium, i.e. to sample the distribution of Eq. 2.24, the detailed balance condition must be satisfied:

$$\begin{aligned} P(...; \boldsymbol{R}_j, \beta_j; ...; \boldsymbol{R}_k, \beta_k; ...) p(j \rightarrow k) = \\ P(...; \boldsymbol{R}_k, \beta_j; ...; \boldsymbol{R}_j, \beta_k; ...) p(j \leftarrow k). \end{aligned} \tag{2.25}$$

If we substitute Eq. 2.24 and 2.22 in Eq. 2.25, we find the following relation for the acceptance probability:

$$\frac{p(j \rightarrow k)}{p(j \leftarrow k)} = e^{\Delta_{j,k}} \tag{2.26}$$

with:

$$\Delta_{j,k} = (\beta_j - \beta_k)(U(\boldsymbol{R}_j) - U(\boldsymbol{R}_k)) \tag{2.27}$$

The condition in Eq. 2.26 can be satisfied by accepting an exchange between two configurations with a Metropolis criterion:

$$p(j \rightarrow k) = \min\{1, e^{\Delta_{j,k}}\} \tag{2.28}$$

If the move is accepted, the configurations are swapped and the velocities rescaled:

$$\begin{aligned} \boldsymbol{R}_j^{'} &= \boldsymbol{R}_k, & \boldsymbol{v}_j^{'} &= \sqrt{\frac{T_j}{T_k}}\boldsymbol{v}_k, \\ \boldsymbol{R}_k^{'} &= \boldsymbol{R}_j, & \boldsymbol{v}_k^{'} &= \sqrt{\frac{T_k}{T_j}}\boldsymbol{v}_j. \end{aligned} \tag{2.29}$$

In PTMETA, metadynamics is performed by each replica, using the same set of CVs at the different temperatures. We indicate with $V_G^{(i)}$ the bias potential acting on the ith replica. With this addition, the total potential felt by the ith replica becomes:

$$U(\boldsymbol{R}_i) + V_G^{(i)}(\boldsymbol{S}(\boldsymbol{R}_i), t), \tag{2.30}$$

and $\Delta_{j,k}$ in the acceptance probability of Eq. 2.28 is modified accordingly:

$$\begin{aligned} \Delta_{j,k} = &(\beta_j - \beta_k)(U(\boldsymbol{R}_j) - U(\boldsymbol{R}_k)) \\ &+ \beta_j[V_G^{(j)}(\boldsymbol{S}(\boldsymbol{R}_j), t) - V_G^{(j)}(\boldsymbol{S}(\boldsymbol{R}_k), t)] \\ &+ \beta_j[V_G^{(k)}(\boldsymbol{S}(\boldsymbol{R}_k), t) - V_G^{(k)}(\boldsymbol{S}(\boldsymbol{R}_j), t)]. \end{aligned} \tag{2.31}$$

The combination of PT with metadynamics is particularly effective since it compensates the individual flaws of the two methods alone. In PTMETA, each system evolves independently while filling its own free-energy profile with Gaussians, until an exchange is attempted. If the move is accepted, the system starts to explore and fill with Gaussians a new region of the CVs space. As a result of this tunneling, different basins can be filled simultaneously and the free-energy difference between them can be evaluated in a shorter time with respect to metadynamics dynamics alone. Metadynamics boosts significantly the sampling efficiency of PT. PT leads the system to cross moderate free-energy barriers on all degrees of freedom, while metadynamics allows to overcome higher barriers on a few selected CVs. The final result is that with PTMETA the effect of neglecting a slow degree of freedom in the choice of the CVs is less significant. The use of this combined method will be seen in Chapter 5 where we have used it in order to understand the folding/unfolding mechanism of β-hairpin.

Chapter 3

The closure mechanism of Cyclin dependent kinase 5

3.1 Summary

In this Chapter I report our simulations on the inactivation mechanism of cyclin dependent kinase 5 by means of metadynamics with path collective variables. Firstly I recall the general problem of kinase inactivation, then I present the technical details of the performed simulations, finally the results are described. The atomistic dynamics of the "open-to-closed" movement of CDK5 has a two-step mechanism: first, the αC-helix rotates by $\sim$45°, allowing the interaction between Glu51 and Arg149; then the CDK5 activation loop refolds to assume the closed conformation. The free-energy profile associated with the global motion has been studied and has led to the identification of a CDK5 intermediate, which could be exploited for drug-design purposes.

3.2 Introduction

Kinases represent a vast protein family involved in vital cellular pathways and responsible for several biochemical functions[5]. Understanding the conformational dynamics of kinases[10] may represent a bridge between their structure and function[15]. Indeed, the

activation of many cellular mechanisms requires kinases to undergo large-scale intrinsic motions, as these proteins can adopt at least two extreme conformations, an "open" state that is maximally active and a "closed" state that shows minimal activity[10]. Structural insights into the open and closed states of kinases have been gained from a number of case studies, where two or more crystal structures of the same protein in different conformations were determined by X-ray experiments[15, 43]. However, an understanding of the energetic profile and the atomistic dynamics of the transition between the open and closed states is still in its infancy. Moreover, knowledge of the structures of possible intermediate conformations sampled during the transition is still very limited. Therefore, X-ray data must be complemented with other approaches able to capture the dynamics of large-scale conformational movements[1]. Recently, conformational transitions have been measured by NMR relaxation experiments with substrate analogues[44, 45, 46, 47] or during catalysis[48, 49] as well as by single-molecule FRET measurements[50, 51, 52, 53, 54, 55]. In this scenario, atomistic simulations could greatly contribute to linking flexibility to function[1, 56, 57, 58]. Although conventional MD studies in explicit water solution can access only a limited time scale (100 ns), large-scale conformational rearrangements occur on the microsecond-to-millisecond time scale. Thus several schemes have been developed to overcome this drawback, most of them based on a coarse-grained representation of the system[59, 60, 61, 62]. This approach can be very useful; however, obtaining a good coarse-grained representation can be difficult, and transferability is a major concern. Here, the use of metadynamics within PCV[39] allow us to investigate kinase plasticity by studying the energetics and dynamics of the intrinsic conformational motions experienced by CDK5 during the "open-to-closed" transition. Cyclin-dependent kinases (CDKs) are key players in the machinery that drives the progression and timing of the eukaryotic cell cycle[63], and malfunctioning of CDKs is the basis of several diseases[64, 65, 66, 67]. To be fully active, CDKs require binding of homologous protein activators known as cyclins and phosphorylation of a conserved threonine of the activation loop. Similarly to other kinases, during the open-to-closed transition, CDKs undergo a large-scale intrinsic rearrangement that results in the rotation of an α-helix by almost 90°and relocation of the activation

loop in a movement of 25 Å[68, 69].

CDK5[70] is a member of the CDK protein family, and it does not seem to be involved in cell-cycle regulation. Moreover, it is selectively activated by p35 or p39, whose expression is limited to neurons and a few other cell types[71]. As a consequence, CDK5 is implicated in neuronal development and maintenance of adult neuronal architecture[72], and its deregulation has been associated with a number of neurodegenerative diseases, such as Alzheimer's disease[73], Parkinson's disease[74], amyotrophic lateral sclerosis[75], Niemann-Pick type C disease[76], and ischemia[77]. Furthermore, CDK5 does not need the phosphorylation of the activation loop to be fully active. Actually, in the X-ray structure of the CDK5/p25 complex[78], the structural requirement for phosphorylation is bypassed by an extended network of interactions between p25 and the CDK5 activation loop. The absence of the phosphorylation step renders CDK5 particularly well-suited for investigating kinase large-scale intrinsic motions, and moreover, it represents a very interesting target for drug discovery.

3.3 Simulation details

3.3.1 Homology modeling

The structures of open CDK5 and the CDK5/p25 complex were obtained from the crystal structure of CDK5 refined at 2.2 Å resolution in a complex with roscovitine (PDB entry 1UNL[79]) by removing the inhibitor and the activator p25 and replacing Asn144 with an Asp, in order to have an active protein. The search for a reactive path starts from the definition of an initial (open) state and a final (closed) state. Since the crystal structure of CDK5 in the closed state is not available, we built a chimeric structure in which the αC-helix and T-loop were obtained by homology modeling of CDK5 on the closed structure of CDK2 (PDB entry 1FIN[80]). These regions show sequence identities (74 and 85% for the αC-helix and the T-loop, respectively) that are even higher than that related to the entire protein (60%). The remaining residues were taken from the original structure of CDK5 in the open conformation (PDB entry 1UNL[79]). The homology model was generated with

the program Modeller version 7.0[81]. The structures thus obtained were ranked on the basis of the internal scoring function of Modeller and validated with PROCHECK[82]. According to the PROCHECK analysis, the homology-built CDK5 closed conformation turned out to be a very high-quality model, with an overall G-factor of 0.04. The Ramachandran plot[83] related to this model showed that more than 90% of the residues were located in the most favored regions.

3.3.2 Molecular dynamics

Preliminary MD simulations of the open state, the closed state as described above, and the CDK5/p25 complex were carried out using the GROMACS 3.2.1 suite[84] and the OPLS-AA force field[85]. Water solvation was described by immersion of the open and closed structures in 13000 TIP3P waters, while for the complex, a larger number (23000) was necessary. Neutrality was ensured by adding one Cl^- to the two isolated proteins and two Na^+ ions to the complex. Periodic boundary conditions and the particle mesh Ewald method to account for long-range electrostatic interactions were used throughout. Bonds involving hydrogens were constrained using the SHAKE algorithm[86], and the time step for all of the simulations was 2 fs. A steepest-descent minimization and thermalization scheme was applied to all the initial structures. The systems were heated from 0 to 100 K in 50 ps and to 300 K in 200 ps, keeping the Cα atoms fixed in their original positions. Next, all constraints were lifted, and the equilibration was continued in the isobaric-isothermal ensemble with Nose-Hoover thermostats[87] and the Parrinello-Rahman barostat[88] for 1 ns. For the open state and CDK5/p25 complex, after the thermalization scheme we performed 10 ns of MD on the canonical ensemble. In these simulations, the root-mean-square deviation (rmsd) of the protein and Cα showed good thermal stability, with a value of 1.5-2 Å.

3.3.3 Building the guess path

After MD thermalization, the open and closed conformations of CDK5 were submitted to the Molmov morphing server[89]. Starting from our initial and final structures, we obtained

five interpolated intermediates. Since the bond lengths are kept fixed to their experimental values and an energy minimization is employed, the resulting sequence of structures is only a rough approximation to the reaction coordinate. In Fig. 3.1, the different conformations of the αC-helix and the T-loop of the five CDK5 intermediate conformers (as output from Molmov) are shown along with the open and closed states of the enzyme.

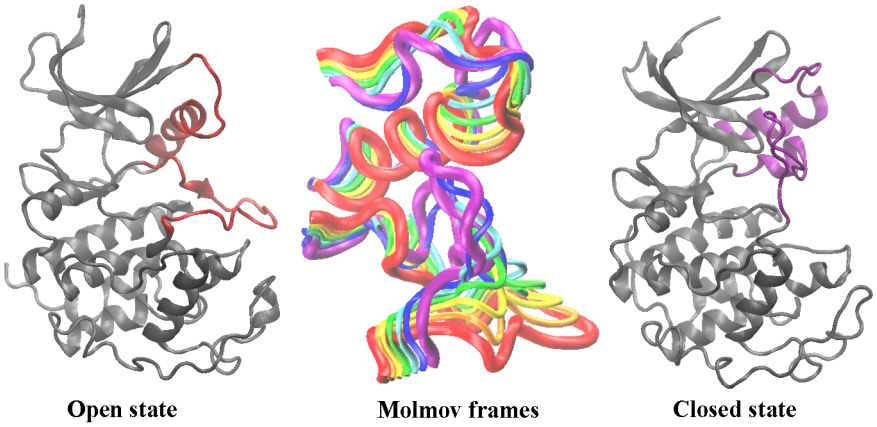

Figure 3.1: Different conformations of the αC-helix and the T-loop of five CDK5 intermediate conformers between open and closed state obtained by means of Molmov interpolation

Each of the five morphing conformations was then hydrated and thermalized at 300 K (see above), and the last snapshots were used for generating the initial guess path. Notably, some morphing conformations were quite stable during the thermalization, while others showed a rapid increase of the rmsd of the Cα atoms. The seven conformations (the open state, the closed state, and the five morphing conformers) were used to generate the guess for subsequent path optimizations and metadynamics-based computations, which were carried out using the path collective variables $s(\boldsymbol{R})$ and $z(\boldsymbol{R})$ described below.

3.3.4 Metadynamics with path collective variables

We used the PCV[39] described in Chapter 2. We defined two variables, $s(\boldsymbol{R})$ and $z(\boldsymbol{R})$, that are able to describe the position of a point in configurational space ($\boldsymbol{R}$) relative to a preassigned path. In the present case, the preassigned path coincides with the morphing

path. The variable $s(\boldsymbol{R})$ provides the progress of the dynamics along the template path, while $z(\boldsymbol{R})$ provides the distance from the template path itself. In the present case, we used the mean square displacement after optimal alignment for calculating the distance square $\|...\|^2$ in Eq. 2.16 and 2.17 in Chapter 2. The choice of the coordinates of atoms to be included in the description is far from trivial: a wrong choice can lead to a loss of performance. An in-depth analysis of several trial trajectories in which we used only the Cα and several N atoms that belong to the αC-helix (30 to 60) and T-loop (145 to 165) showed that the side chains of Lys33, Arg50, Glu51, Asp144, and Arg149 also must be taken into account to effectively simulate the overall movement. For the discrete representation to be meaningful, we chose the nodal points l to be as equidistant as possible and the value of λ to be comparable to the inverse of the mean square displacement between successive frames. We then ran a metadynamics[3] simulation to reconstruct the FES as a function of s and z. We used metadynamics in the direct version[90], as implemented in NAMD[91]. The height of the Gaussians was set to 0.3 kcal/mol and the addition frequency to 1 ps. The width of the Gaussians was not fixed but was determined by the fluctuations in s and z measured over a time interval of 2 ps. In order to avoid integration problems with too-narrow Gaussians, a lower-bound limit of 0.03 in the appropriate unit was set for each coordinate. We first optimized the path with $P = 7$ using the procedure described in ref. [92]. It was clear, however, that this discretization was not sufficient to describe the transition with the necessary resolution. We thus increased P to 18 by selecting additional intermediate configurations from among the small-z configurations obtained during the metadynamics. We chose these configurations such that they were equally spaced with a distance of 1.31 Å. The 18 frames were used to define a reference path using $\lambda = 1.33$ Å^{-2} in Eq. 2.16 and 2.17. The path was not further improved, but it was sufficiently accurate for the study of FES(s,z) and provided us with reliable information on the open-to-closed transition and its associated free-energy profile.

3.4 Results

The structural organization of CDK5 is reported in Fig. 3.2 and hereafter briefly summarized. CDK5 exhibits the classical bilobal kinase fold, where the N-terminal domain (N-lobe) is composed mainly of the β-sheet and one α-helix (the αC-helix) (Fig. 3.2a).

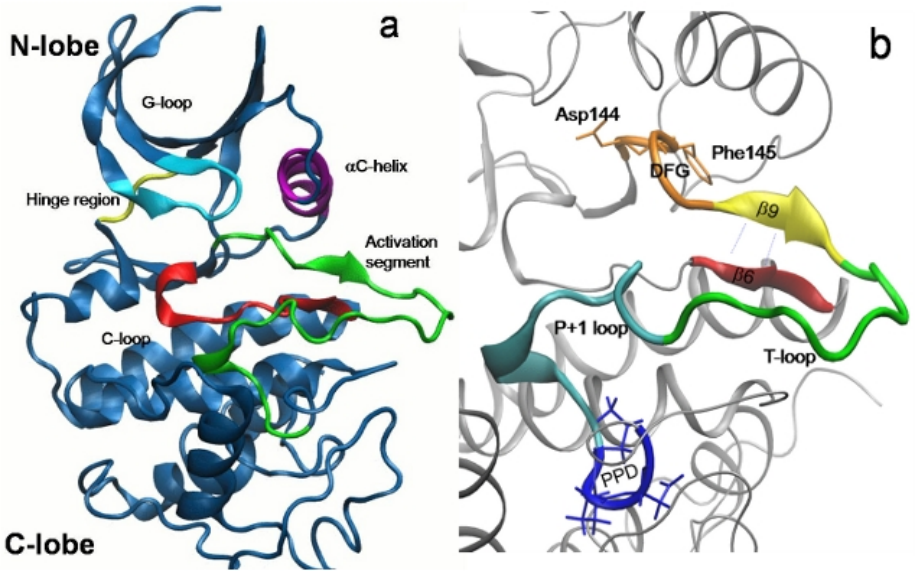

Figure 3.2: Three-dimensional structure of CDK5 in the open state. (a) Classical bilobal structure of kinase: (cyan) the G-loop (aa 11-18); (violet) the αC-helix (44-57); (yellow) the hinge region between the N- and C-lobes of the protein (81-84); (red) the C-loop (122-131); (green) the activation segment (144-171). (b) Details of the activation segment: (orange) the DFG-in conformation, in which the Asp144 is pointing inside the cleft; the interaction between (yellow) β9 and (red) β6; (green) the T-loop; (cyan) the P+1 loop; (blue) PPD, the C-terminal part of the activation segment.

The C-lobe is predominantly α-helical and is linked to the N-terminal by a flexible hinge (81 to 84 aa). Between the lobes there is a deep cleft that natively binds ATP. The ceiling of the ATP-binding site, named the G-loop (11 to 18 aa), is quite flexible, being rich in glycines[93]. Two structural elements are involved in the interaction with the activator[78] (p25) and in the catalytic function: the αC-helix, whose sequence is well-conserved in the CDK family, and the activation loop (T-loop), which binds the two

terminal lobes (145 to 165 aa). The latter belongs to the activation segment, which is shown in detail in Fig. 3.2b. This is a structurally well-characterized portion of kinases that begins with the conserved DFG tripeptide (144-146 following the CDK5 numbering) and comprises the magnesium-binding loop (143-148), β9 (149-150), the activation loop (T-loop in CDKs) (151-160), and the P+1 loop (161-169), ending with a second conserved tripeptide motif (APE in general but PPD in CDK5 (170-172)). In the C-lobe there is also another fundamental loop, called the catalytic loop (C-loop, Fig.3.2a), that is involved in the reaction catalysis (122-131 aa).

3.4.1 Closure movement of CDK5: a two-step mechanism

In a preliminary calculation, we assessed the effect of p25 on the stability of the open form of CDK5. In Fig. 3.3, the rmsd during a 10 ns MD simulation is reported for the CDK5/p25 complex (Fig. 3.3 A) and for CDK5 alone (Fig. 3.3 B).

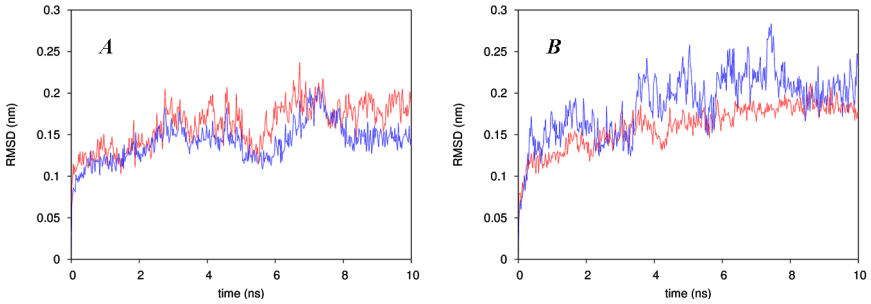

Figure 3.3: RMSD calculated on all the Cα of CDK5 (red plot line) and on the Cα of the αC-helix and the T-loop (residues 35-55 and 145-165, blue plot line). A: CDK5/p25 complex. B: CDK5 alone.

It can be seen that CDK5 in the complex with its activator p25 (Fig. 3.3 A) is more rigid than CDK5 alone (Fig. 3.3 B) and that most of the flexibility of uncomplexed CDK5 comes from the αC-helix and the T-loop (blue trace in Fig. 3.3 B). The stabilizing effect of the activator suggests that its role is to favor the open structure. This conclusion is strengthened by the observed dependence on activator concentration[94]. We simulated the closure process in the absence of the activator, which is a necessary step toward a

full understanding of the enzyme functioning. To determine the closing path (see Simulation details), we used metadynamics together with the method of Branduardi et al.[39]. Although the initial guess path was constructed using a standard bioinformatics tool[89], the final optimized path was very different from the initial guessed one.

Our simulations show that the process takes place in two steps: in the first step, the αC-helix rotates by 45°(Fig. 3.4 A, B), and in the second step, the T-loop closes the catalytic cleft and the αC-helix completes its rotation (Fig. 3.4 C, D). The first event leading to the rotation of the αC-helix is the breaking of the salt bridge between Lys33 and Glu51, which plays a key role in the kinase activation mechanism[95] (Fig. 3.4 A, B). The role of this salt bridge is to anchor the αC-helix in its open position. Solvation and the formation of a new salt bridge with Arg149, which is part of the activation segment, enable a partial rotation of the αC-helix. In the open conformation, Lys33 and Glu51 interact strongly and are assisted in stabilizing the αC-helix at its position by the close proximity of Asp144. The breaking of this bond takes place in a concerted fashion. Arg149 approaches Glu51, thus weakening the Lys33-Glu51 bond; this leads to a bond switch between Lys33-Glu51 and Lys33-Asp144. The end result is that the Lys33-Glu51 bond is disrupted and a new salt bridge, Glu51-Arg149, is formed, allowing the partial rotation of the αC-helix. The DFG motif did not experience any in-out conformational movement, as observed in the crystal structures of CDK2 and CDK7 but at variance with CDK6, which can adopt a DFG-out conformation when it is bound to the inhibitor Ink4[14]. In the second step, the αC-helix completes its rotation while the T-loop refolds to assume the final closed conformation. During the entire process, Arg149 and Glu51 remain bound (Fig. 3.4 C,D). Also in this second stage, Arg149 plays a major role by helping the T-loop to rearrange and maintain a connection between the C- and N-lobes. The mechanism that we have found by optimizing the path is closely related to one of the possible mechanisms described in ref [60] for the homologous Src kinase. In that paper, the authors used coarse-grained molecular representations and found that the closed-to-open transition can occur in two possible ways. The first pathway is similar to what we found in CDK5 and implies the opening of the activation loop followed by rotation of the αC-helix. Also, the switching of

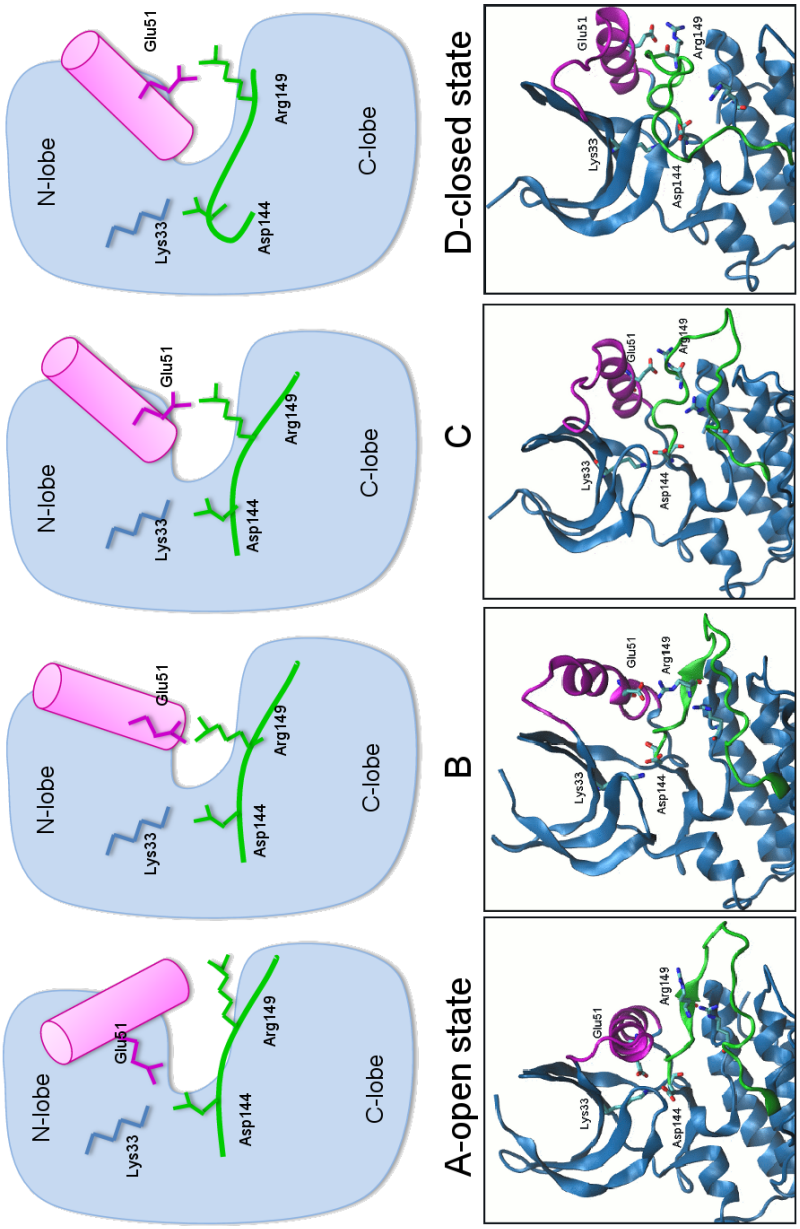

Figure 3.4: Different conformations assumed by CDK5 in going from the open state to the closed state: (upper panel) schematic illustration of the transition; (lower panel) corresponding cartoons of CDK5. From A to D, the figures show the progressive disruption of the Glu51-Lys33 interaction, the formation of the salt bridge between Arg149 and Glu51, the rotation of the αC-helix (pink), and the movement of the T-loop (green).

an electrostatic network, in particular between the highly conserved residues K295-E310 and E310-R409, plays a fundamental role in their case, as previously noted in ref [96] and recently confirmed in ref [97]. On the contrary, we did not find evidence for the partial unfolding mechanism described in ref [60]. An added bonus of our atomically detailed pathway is that it shows the importance of hydration/dehydration of key residues during the restructuring of T-loop (Fig. 3.5).

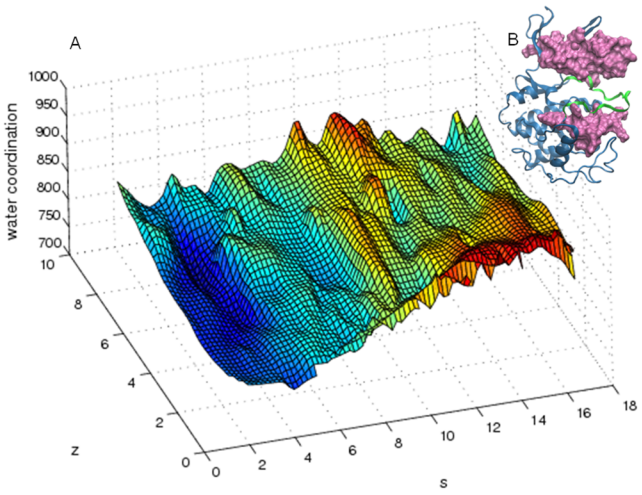

Figure 3.5: A: Water coordination of T-loop as a function of s and z. B: The water coordination was calculated on the oxygen atom of the residues 10-20, 30-55, 120-125, 166-181 (shown in pink in B),145-165 (shown in green in B).

Initially, the loop is relatively dehydrated. Halfway through the transition, the T-loop hydration reaches its maximum, being exposed to the solvent. At the end, water molecules must be expelled from the cavity, and T-loop hydration is reduced. After closure, not only have the αC-helix and T-loop moved, but also the G-loop is displaced, making ATP binding impossible[95]. In the open structure, the initial segment of the T-loop forms a β-sheet motif (β9), while in the final closed structure, it is disrupted and refolds into a coil. This is at variance with the result of the homology modeling, which predicted the formation of an α-helix, but consistent with several experiments on different members of the kinase family, which failed to resolve this structural element[98, 99] in the closed state.

3.5 Free-energy surface and metastable intermediate state

The use of metadynamics with the PCV also permits the reconstruction of the FES along the optimized reaction coordinate. In Fig. 3.6, the bidimensional FES for the open-to-closed switching movement of CDK5 is reported.

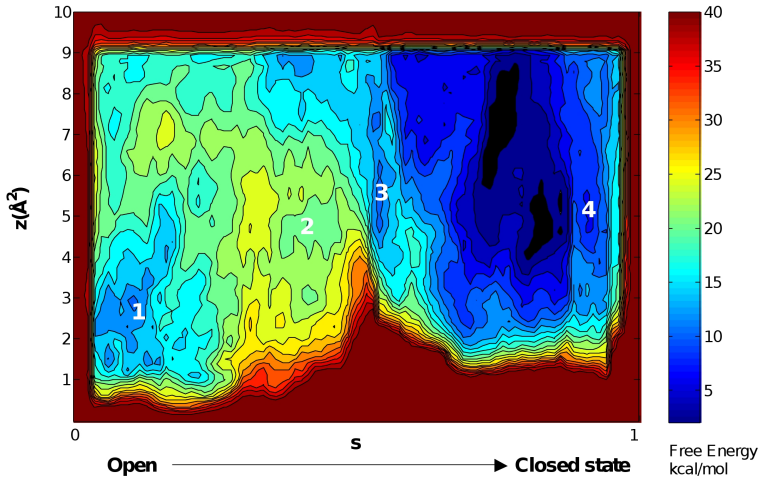

Figure 3.6: Free-energy surface reconstructed as a function of s and z. The energy separation between contours is 2 kcal/mol. The numbers 1-4 correspond to the states described in Fig. 3.4.

The FES, obtained from a 150 ns metadynamics simulation, was reconstructed as a function of the progression along the path (s) and the distance from the path itself (z). The projection of the FES on the s and z variables is an unusual one and probably requires a more detailed description. The z variable describes the distance from the path in the MD simulation. This has important consequences, as some energy basins that seem to be far away from the pathway are not "distinct" structures/mechanisms but indicate substantial entropic contributions to those basins. This is evident in the proximity of the closed state and is due to the intrinsic fluctuations of the activation loop. This observation is also in agreement with the findings of ref [60] that describe a very flexible activation loop in the case of the Src kinase. Additional simulations in which the overall movement was

split into two subpaths were performed, and they quantitatively confirmed the FES and mechanism obtained. As expected, the open and closed states (A and D, respectively, in Fig. 3.4) correspond to two deep free-energy basins. The closed state was 4-6 kcal/mol lower in free energy than the open one, so in the absence of an activator, the protein remains closed. The activation barrier is predicted to be 16-20 kcal/mol. This high value was confirmed by repeated metadynamics simulations both with path-like variables and with a different set of CVs. In order to check these results and also to confirm the FES obtained and the mechanism, we studied the two steps of the closure movement, into two separate simulations splitting the overall movement into two sub-paths. The first sub-path corresponds to the transition from the open state to the intermediate B of Fig. 3.4, the second sub-path is related to the movement from intermediate B of Fig. 3.4 to the closed state. The first sub-path, in which the main event is the disruption of the Lys33 and Glu51 salt bridge and the formation of the Glu51 and Arg149 interaction, was studied by means of two independent metadynamics runs: in the first one, we used two distances (Lys33-Glu51 and Glu51-Arg149) as CV, in the second the path collective variable (only s and z restrained to a value around zero) and the water coordination number for two residues (Arg50 and Glu51). The water coordination number was calculated by means of the "coordination number" of all oxygen of the water and oxygen of Arg50 and Glu51:

$$C = \sum_{i \in W} \sum_{j \in P} \frac{1 - \left(\frac{r_{ij}}{r0}\right)^8}{1 - \left(\frac{r_{ij}}{r0}\right)^{14}}. \tag{3.1}$$

where W is the index of water molecules, P represents the aminoacids Glu51 and Arg50, r_{ij} is the distance between oxygen atoms i of the water and j of the protein, $r_0 = 5.5$ Å is a parameter that takes into account their typical distance. The two independent simulations agree on a difference of about 5 kcal/mol between the open and the metastable state B. The second sub-path describes the final part of the transition: in order to simulate it we first extracted a path from previous trajectories, optimized it by means of the PCV method, and used it as a path for the metadynamics simulations. The extracted frames were 28; after optimization, we used them for metadynamics simulation with $\lambda = 5.86$ Å^{-2}.

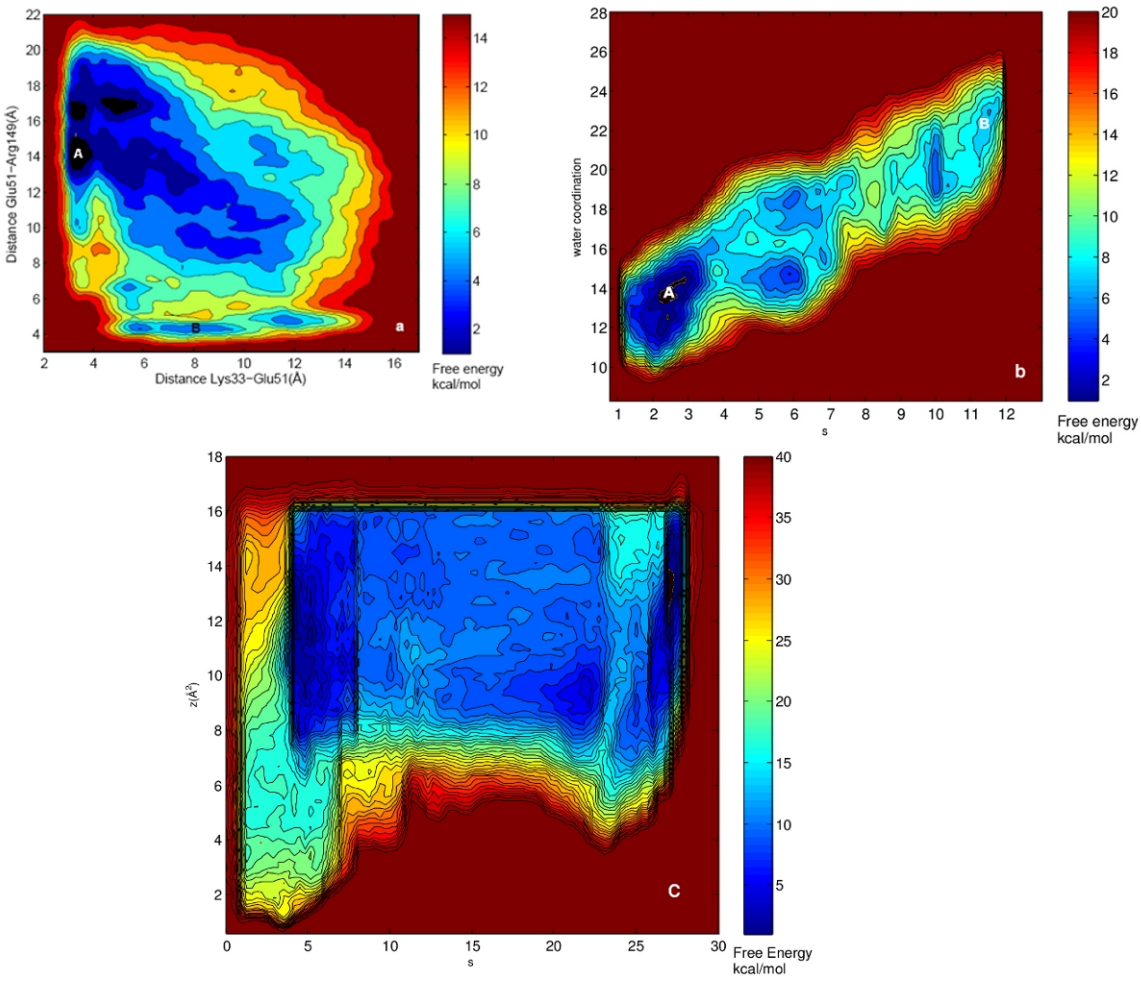

Figure 3.7: A: free energy profile as a function of Lys33-Glu51 and Glu51-Arg149 distances (the isoline separation is 1.0 kcal/mol). b: free-energy profile as a function of s and water's coordination of Arg50 and Glu51 (the isoline separation is 1.0 kcal/mol). c: free energy profile as a function of s and z of the subpath (the isoline separation is 1.0 kcal/mol).

The free energy landscape for the latter sub-path is reported in Fig. 3.7. As in the previous FES, also this free energy is largely in agreement with that of Fig. 3.6.

Although we cannot rule out an error due to the force field or the existence of different pathways, the similarity of the path found here to that previously described for the homologous Src kinase, in addition to biological considerations, makes us believe that the high barrier is indeed physical. This result is consistent with a mechanism in which p25, p35, and p39 have a fundamental catalytic role in the activation process, in agreement with previous suggestions[94, 100, 101]. We report in 3.4 the conformations corresponding to the minima observed in the FES that are most relevant in the transition from the open to the closed state (labeled as points 1, 2, 3, and 4 in Fig. 3.6). The main geometric features of these conformations are reported in Table 3.1.

Table 3.1: Distances (Å) between the main residues involved in the transition from the open state to the closed state

	open state	**intermediate state**	**closed state**
Phe145–Leu54	5.34	9.40	13.59
Phe151–Asn121	1.95	9.39	11.73
Arg149–Leu123(CO-HN)	2.10	6.69	8.36
Arg149–Leu123(NH-OC)	2.34	5.53	6.61
Lys33–Asp144	3.39	3.46	5.28
Lys33–Glu51	3.21	14.80	16.11
Glu51–Arg149	15.32	3.86	3.95
Glu51–Arg50	11.34	9.05	6.05
Thr164–Asp126(HRD)	3.76	4.52	1.82

The deepest intermediate minimum (point 3 in Fig. 3.6) is a metastable state whose dynamical relevance was checked in an unconstrained MD run of 10 ns. During this time, the intermediate state did not exhibit any significant rearrangement. This intermediate shows structural features that differ from those of the open and closed states. The αC-helix is rotated by 90°, and while the Lys33-Arg149 salt bridge is formed as in the closed

state, the T-loop is in an open conformation with the β9 sheet already refolded in a coil structure. In the Src and Hck kinases[102, 103], an intermediate has been crystallized in which the αC-helix is also rotated by 90°, but the structure otherwise differs from ours. In particular, in the Src and Hck intermediates, Glu310 in the αC-helix interacts with Arg385 and not Arg409, which corresponds to Arg149 in CDK5. If the existence of such an intermediate were to be confirmed, it appears that enough elements of diversity exist between the intermediate and the open state of CDK5 for new drugs that target the intermediate to be designed. The existence of an intermediate state proved to be crucial in the case of Bcr-Abl tyrosine kinase[15, 104], for which it has been shown that the anticancer drug imatinib selectively binds to the intermediate.

3.6 Conclusions

The intrinsic motions of CDK5 have been investigated here by means of atomistic simulations. In 150 ns of metadynamics, we have sampled the CDK5 conformational rearrangement underlying the transition between its open and closed states. This process turned out to be rather complex, and it takes place into two steps. First, the salt bridge between Lys33 and Glu51 is broken, leading to 45° rotation of the αC-helix and the formation of a salt bridge between Glu51 and Arg149. Second, a highly concerted motion of the αC-helix and the T-loop leads to the closed conformation, with the αC-helix rotated by 90° and Arg149 and Glu51 form a salt bridge and are exposed to the solvent. During this step, we could detect a complete refold of the T-loop leading to the CDK5 closed conformation. In addition, we discovered the key role of Arg149 during the transition, as it works as a linker between the N- and C-lobes, allowing a concerted movement of the two domains.
The free energy of the process was accurately estimated, pointing to the closed state being more stable than the open one by 4-6 kcal/mol. Furthermore, a previously undisclosed CDK5 low-energy intermediate conformation was identified and its stability assessed. A similar intermediate conformation was previously crystallized in two related protein kinases, Src and Hck[102, 103], giving support to our finding. This result, together

with the similarity of the pathway found here to that described in the homologous Src kinase[60, 96, 97], gives us confidence that even if what we propose is a model, its relevance might be verified by experiments such as those reported in ref [105]. While understanding how the dynamics of kinases influences their function is a great challenge for basic science, the determination of the open-to-closed atomistic path and the associated free-energy profile may provide undisclosed kinase intermediate conformations suitable for addressing the structure-based design of potent and selective kinase inhibitors.

Chapter 4

P38 dynamics in free and ligand-bound form

4.1 Summary

In this chapter I present the study of the inactivation mechanism of p38α MAP kinase. In this protein a typical conformational change, called "DFG" flip characterizes the movement of three well conserved aminoacids at the beginning of the activation loop. During this movement the Aspartate moves away from the ATP binding site, and its position is taken by the neighboring Phenilalanine. This transition allows the creation of an additional binding site that can be targeted by some selective inhibitors.
In this Chapter the DFG-in/out transition for the apo and a ligated forms of the enzyme has been studied by means of metadynamics simulations. The data coming from X-ray, NMR and kinetic binding assays are complemented revealing the main molecular features of the DFG-in/out motion.

4.2 Introduction

P38 kinase plays a fundamental role in regulating the production of pro-inflammatory cytokines. Thus, blocking this kinase may offer an effective therapy for treating many in-

flammatory diseases[106, 107]. However, despite recent successes in kinase inhibitor drug discovery, several important challenges are still open. In particular the search for potent and selective inhibitors and the understanding of kinase drug resistance represent key issues[7]. Most kinase inhibitors are ATP-competitive molecules, which interacts with the kinase in a typical conformation called DFG-in, defined as type I inhibitors (see Fig. 4.1). In this state the regulatory activation loop is open and extended, which allows ATP and substrates to bind. The DFG motif, at the beginning of the activation loop presents the Asp side chain oriented towards the ATP binding site. Given the highly conserved nature of ATP binding site, type I inhibitors present severe selectivity limitations. Thus a modern strategy to find new kinase inhibitors is trying to identify new chemical entities that bind to and stabilize the catalitically incompetent "DFG-out" state. This conformation results from structural changes of the activation loop induced by the flip of the highly conserved DFG motif. Such movement allows the formation of a less-conserved allosteric binding site adjacent to ATP binding site. This site can be occupied by different inhibitors, called type II. These inhibitors also bind through an hydrogen bond to the backbone at the hinge region occupying in this way the ATP-binding site. Type II inhibitors have been reported to be more potent and to increase the drug-target residence time, thus improving the therapeutic effect. The pyrazolourea BIRB-796[108], a selective inhibitor of p38, developed by Boehringer-Ingelheim, and imatinib (Gleevec)[104, 109], an Abl/cKit/PDGFR inhibitor, are well-known examples of type II inhibitors[110, 111, 112].

Additionally, a number of high affinity inhibitors that bind exclusively to the allosteric site of p38α are referred to as type III inhibitors. Such compounds also stabilize the DFG-out conformation, without interacting with the ATP binding site. They are usually studied as starting leads for the design and synthesis of type II inhibitors. Because these molecules that bind to this less-conserved site of the protein have good selectivity profiles and improved pharmacological properties, a growing attention is being put in methods that could describe the movement from DFG-in to DFG-out state or that can discriminate which is the preferred state a ligand is bound to. To this purpose, p38 kinase has been extensively studied both experimentally and theoretically. Several structures of various inhibitors in

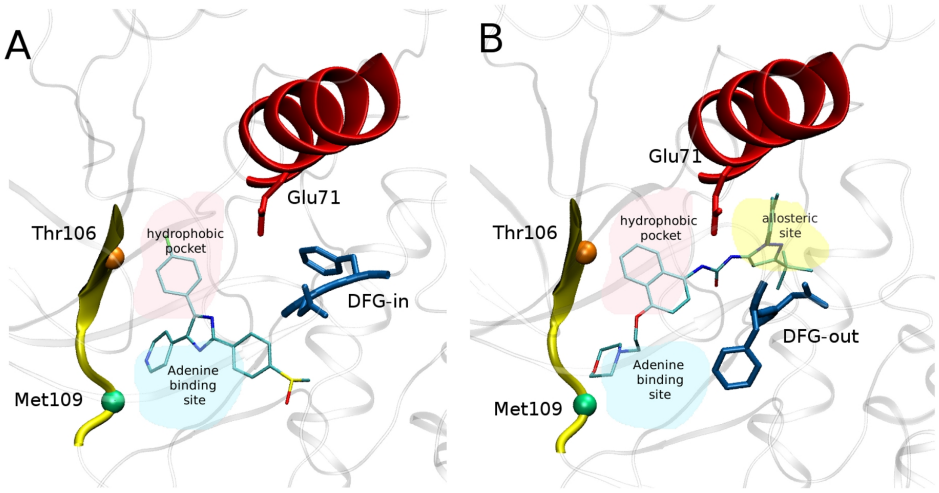

Figure 4.1: Type inhibitor I and II inside p38 kinase. In panel A the typical binding mode for SB-203580, a type I inhibitor, is represented. Panel B shows the interactions between BIRB-796, a type II inhibitor bound with the protein that assumes a DFG-out conformation. The hinge region of the protein is in yellow, the green sphere represents Met109, that interacts with both ligands through an hydrogen bonds, the orange one shows the position of Thr106 that makes hydrophobic contacts with both the inhibitors. In red the position of αC-helix with Glu71 is represented. It can be noted that the type II inhibitors interacts both with the nucleotide binding site and extends into the adjacent allosteric site.

complex with this protein have been resolved providing useful information on binding modes and inhibition mechanisms, helping also to rationalize the structure-activity relationships of many ligands. P38 kinase has been crystallized in both DFG-in and DFG-out state. However, while up to now the apo form state has been resolved only in the DFG-in state, the ligated form has been found in the DFG-in and DFG-out states. In the majority of the deposited structures, the whole loop has never been solved, indicating that this region is very flexible. Because of the lack of molecular information about the dynamical activation process of p38 the mechanism of the DFG flip can only be drawn schematically. Important information in this direction can be obtained by NMR[113] and inhibitor ki-

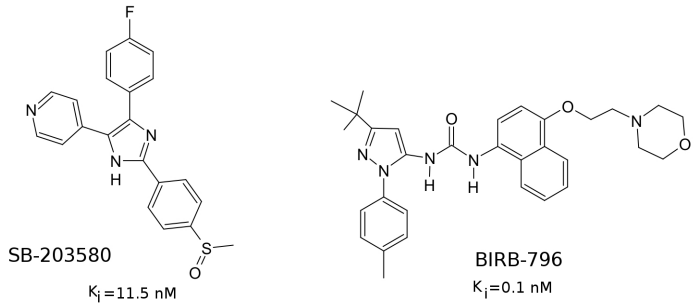

Figure 4.2: Chemical structures of a type I, SB-203580, and a type II, BIRB-796 inhibitors. Data for compounds are taken from ref. [113].

netic studies[108]. NMR spectroscopy studies have shown that there is a conformational equilibrium between DFG-in and DFG-out in p38 which varies in response to the different inhibitor bound. The apo form of the protein has a conformational DFG-in/out equilibrium that happens at intermediate timescale which can only marginally be resolved by NMR. However a number of information have been gathered. The binding of type II inhibitors, like BIRB-796[108], freezes the protein in the DFG-out conformation, while the addition of the type I inhibitor, like SB-203580[114], does not prevent a DFG-in/out equilibrium. In line with this NMR experiments, both conformations of the DFG-motif have been found ligated by this inhibitor which however assumes two different binding modes in the DFG-in and DFG-out conformation. Like SB-203580 other type I inhibitors seem to bind the protein either in DFG-in or DFG-out conformation[113, 115]. Indirect evidences of the DFG-in/out equilibrium come from the analysis of the binding rates of different ligands. In fact, this kinetic experiments shows that BIRB-796 binds 200-fold slower than a comparable type I inhibitor[108, 116]. Similar results have been obtained also by means of other experimental techniques such as fluorescence kinetics assays[115]. Given these data one can hypothesize that the DFG motif in apo p38 is in a slow conformational equilibrium between the two states, the presence of type II inhibitors renders this conformational equilibrium extremely slow, while type I inhibitors interfere with such equilibrium to a smaller extent.

From a theoretically point of view, given the long time scale associated with this movement, simulations of hundreds of ns of standard molecular dynamics do not allow a simulation of the entire DFG-in/out transition. Different strategies to overcome this hurdle have been proposed for p38: in ref. [117] the authors uses high temperature molecular dynamics in order to accelerate the sampling of the transition, in ref. [118] the use of accelerated molecular dynamics has been applied. Another study of the DFG transition has been carried out on a kinase different from p38 by means of targeted molecular dynamics[15], while, Shan et al.[119] have sampled the transition in a very long (micro-seconds) molecular dynamics runs. However they could only observe the transition by induce a key mutation. Although such studies are really important and provide insights in the DFG-in/out mechanism, they cannot describe exhaustively the whole dynamical character of the DFG transition and the free energy profile associated with this movement. Moreover the role of different inhibitors on such motion has not up to now been reported.
Here we use well-tempered metadynamics simulations to enhance the sampling of the DFG-in/out transition at atomistic level and to obtain the free energy associated to the movement. As collective variables we used the path collective variables. In the apo form the free energy profile associated with this transition confirms the existence of an equilibrium between the two states. Moreover an intermediate state in between the ending points has been found and it will be used for future drug design study. Once a type I inhibitor, like SB-203580 is bound to the protein it does not interfere too much with the DFG-in/out equilibrium. From these results the dynamical aspects of DFG-in/out movement are disclosed and the data obtained here will be of precious help for future experiments and structure-based drug design.

4.3 Simulation details

4.3.1 Molecular dynamics

The X-ray structure of p38α in the DFG-in state was taken from 1A9U.pdb[114]. The starting configuration for the apo state was obtained by removing the inhibitor SB-203580,

while it was retained in the simulation of ligand bound to the protein. In this structure the protein is completely solved and does not present any mutations, moreover its structural characteristics are very close to the apo structure of mouse p38α protein (1P38.pdb). For the DFG-out state, since there is not an apo structure of the protein in the DFG-out state, we used the coordinates of 1W82.pdb in which the protein is crystallized together with a type III inhibitor. The three structures, APO-in, APO-out, SB-203580 with DFG-in conformation were equilibrated with a 5-ns molecular dynamics simulation under NPT conditions at 1 atm and 300 K, after proper minimization. We used AMBER99SB[120] force field and NAMD2.7[91] molecular dynamics simulation code. For what concerns the parametrization of the ligand, its charges were computed using the restrained electrostatic potential (RESP) fitting procedure[121]. First the ESPs were calculated by means of the Gaussian package[122] using a 6-31G* basis set at Hartree-Fock level of theory and then the RESP charges were obtained by a two-stage fitting procedure[123]. All the simulations were carried out in explicit solvent, using TIP3P water model. Periodic boundary conditions and the particle mesh Ewald method to account for long-range electrostatic interactions were used throughout. Bonds involving hydrogens were constrained and the time step for all of the simulations was 2 fs.

4.3.2 Metadynamics and path collective variables

Building on our previous experience on the closure movement of CDK5 (see Chapter 3), we used path collective variables in order to describe the DFG-in to DFG-out transition. As state A we chose the X-ray conformation of the protein in the DFG-in state (1A9U.pdb), while, for state B, we used the protein in the DFG-out conformation. The transition from A to B, in this case, involves the rotation of about 180° of the phenylalanine in the DFG motif and the movement of the aspartate that moves out of the ATP binding site. The flip of these two amino acids is coupled with the movement of the activation loop. Several X-ray structures show in fact that also this part of the protein is associated with the transition from the in to the out DFG state. We described the transition between the DFG-in to DFG-out state using five frames, $\boldsymbol{S}(1)...\boldsymbol{S}(5)$, with $\boldsymbol{S}(1)$ and $\boldsymbol{S}(5)$ being the

DFG-in (state A) and the DFG-out (state B), respectively. The three intermediate frames have been obtained from a linear interpolation between the two ending structures by means of g_morph tool of GROMACS4.0[124]. At variance with CDK5 closure mechanism, here we used CMAP as a reduced representation of $S(\boldsymbol{R})$. We have measured the square distance $\|...\|^2$ of a generic state from a point belonging to the reference path using this formula:

$$\|\boldsymbol{S}_C(\boldsymbol{R}) - \boldsymbol{S}_C(\boldsymbol{l})\|^2 = \sum_{j>i} (\boldsymbol{C}_{i,j}(\boldsymbol{R}) - \boldsymbol{C}_{i,j}(\boldsymbol{l}))^2 \tag{4.1}$$

where the nearest neighbors are excluded from the sum. In this case $\boldsymbol{S}(\boldsymbol{R})$ is given by the $j > i$ elements of the contact map matrix ($\boldsymbol{C}(\boldsymbol{R})$) defined as:

$$\boldsymbol{C}(\boldsymbol{R}_{i,j}) = \frac{1 - (\frac{r_{i,j}}{r_0})^p}{1 - (\frac{r_{i,j}}{r_0})^q} \tag{4.2}$$

where $r_{i,j}$ is the distance between the ith and jth couple of atoms within the protein. Proper values of of p, q, r_0 are to be set. The choice of the contacts to include in the representation of the transition is crucial because they have to be representative of the movement under investigation. After several trials, the used contacts are those listed in the following table (4.1): since the DFG flip involves conformational changes of the activation loop beyond the DFG motif itself, several contacts within the activation loop have been introduced in the contact list.

We used the same reference path for both the transitions studied. The λ value was set to 1.22. We then ran well-tempered metadynamics[4] simulation to reconstruct the FES as a function of s and z. We ran two independent simulations each of about 200 ns: one for the apo protein, one for the protein with SB-203580. Well-tempered metadynamics parameters were the following: the bias factor was set at 25, a Gaussian function was added every 1 ps with 0.04 and 0.2 width for s and z variables. The initial Gaussian height was set to $0.3 * T/T0$ kcal/mol, where $T0 = 300K$.

Table 4.1: Contacts included in the definition of $S(\boldsymbol{R})$ and $Z(\boldsymbol{R})$ variables

Contact number	Atom i	Atom j	R_0	p	q
1	Phe169Cβ	Ile84Cα	10.3	40	80
2	Phe169Cβ	Leu75Cα	10.5	30	50
3	Phe169Cβ	Met78Cα	13.0	20	40
4	Phe169Cβ	Ile141Cα	13.0	14	28
5	Phe169Cβ	Ile146Cα	13.0	12	26
6	Phe169Cβ	Ile147Cα	11.6	8	20
7	Phe169Cβ	Met109Cα	15.5	50	90
8	Phe169Cβ	Ala157Cα	12.8	30	60
9	Phe169Cβ	Ser154Cα	10.6	16	32
10	Phe169Cβ	Asp112Cα	14.0	30	44
11	Asp169Cβ	Asn155Cα	6.7	30	60
12	Asp169Cβ	Ala144Cα	16.0	40	80
13	His148Hϵ	Asp168O	4.8	12	24
14	Asp176O	Met179H	3.8	6	12
15	Asp177O	Thr180H	2.9	16	40
16	Glu178Cα	His174Cα	7.0	26	50
17	Leu168Cβ	Ile147Cβ	12.6	6	14
18	Leu168Cβ	Ile147Cβ	12.6	6	14
19	Leu168Cα	Leu74Cα	12.0	18	40
20	Leu168Cα	Ile146Cα	14.0	10	24
21	His174Nϵ	Asp150Cγ	8.9	6	16
22	His174Nϵ	Arg189Cθ	9.1	12	26
23	Arg149Cα	Glu170Cα	9.0	8	20
24	Ala172Cα	Ile147Cα	14.0	22	50
25	Arg173Cθ	Ile146Cα	19.0	26	44
26	Thr175Cα	Ile146Cα	18.5	80	99
27	His174Cα	Asp150Cα	10.0	40	90

4.4 Results

4.4.1 DFG-in/out transition in the apo form of p38

The DFG-in/out flip mechanism in p38

P38 kinase shows the classical bilobal tridimensional structure of the kinases. In the DFG-in conformation the Asp168 points into the ATP-binding site and is bound with a salt bridge to Lys53. This residue is also in contact with Glu71, being the αC-helix in the Glu-in position conformation. Phe169 is localized into a hydrophobic cleft making contacts with the hydrophobic side chains of Ile84, Leu74, Val83, Ile146, Ile141. On the other hand, in the DFG-out state, the DFG motif adopts a completely different conformation in which the Asp and Phe side chains exchange their positions, the first leaving and the latter occupying the ATP-binding site. This conformational change can be described by a rotation of about 180° of these key residues. A careful visual inspection of the atoms positions of the activation loop in both the DFG states, indicates that the flip might also involve conformational changes in the rest of the loop. In order to use PCV, after having defined the initial and the final states, one has to find an initial guess path. To do that we used g_morph, a tool of Gromacs software package[124], that creates, using linear interpolation, intermediate conformations given an initial and a final state. Using this tool, three interpolated intermediate frames were generated. The obtained frames represent the reference path for PCV used in the metadynamics simulations. As discussed in the Simulation details (4.3), the choice of the contacts to be included in the contact map matrix is far from trivial. After a number of attempts, we found the final setting of the contact map list which include several contacts that describe the movement of Asparate and Phenilalanine side chain. Moreover, given the high flexibility of the activation loop, we added other contacts to the list that take into account its movement (see Table 4.1).

In the apo form, starting from the DFG-in conformation we found a “clockwise” rotation (see Fig. 4.3) mechanism of Asp and Phe around the loop axis of about 180°. The motion of the two residues is concerted: the Asp side chain passes over the backbone of the DFG motif, while the Phe moves underneath. The Asp progressively rotates of about 90°, while

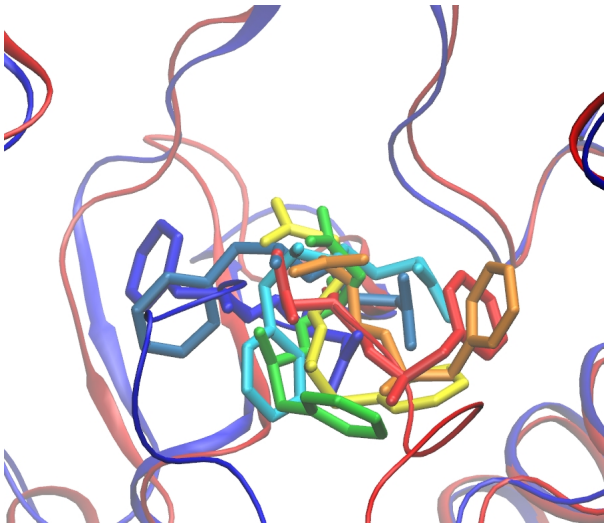

Figure 4.3: DFG-in to out transition in apo form of p38 kinase: snapshots are colored on a rainbow scale with red and blue indicating the DFG-in and DFG-out conformations respectively. It can be seen the clockwise rotation of Asp168 respect to the Phe169

Phe stacks under the activation loop being involved in close contacts with Leu171 and Arg149. Finally, in the DFG-out state also the Phe side chain reaches its final location interacting with Leu167. The rotation of Asp and Phe induces also a global movement of the activation loop of about 7Å passing from the DFG-in to DFG-out state. Along the transition the αC-helix does not experience any important conformational movement such those observed in other kinases, like Abl or Src[15].
Finally the G-loop, in the N-lobe of the protein, shows a great flexibility along the transition forming close contacts with the Phe side chain.

The Free Energy Surface and the presence of an Intermediate State

In Fig. 4.4 the free energy surface for the DFG-in/out movement of p38 kinase as a function of s and z is reported.

This profile has been obtained from 200 ns of metadynamics simulation. The two collective variables describe the progression along the used path (s) and the distance from the path itself (z). The presence of several recrossing events during the metadynamics

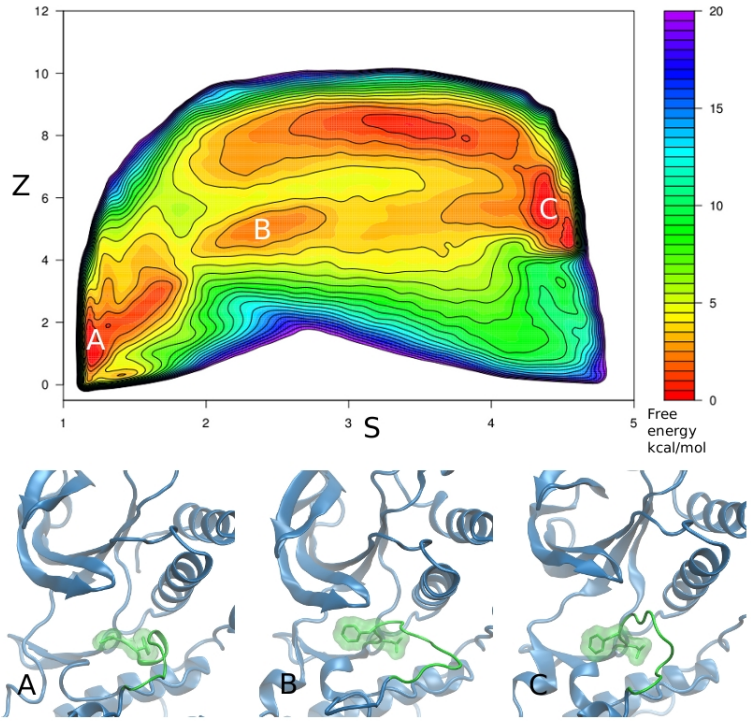

Figure 4.4: Upper panel: free energy surface as a function of s and s variables of apo p38 kinase. Contour lines are plotted every 1 kcal/mol, color legend is in kcal/mol. Lower panel: conformations of the protein corresponding to minima A, B, C respectively.

simulation and the use of the well-tempered formalism reassure us on the convergence of the free energy profile. In Fig. 4.4 the three relevant minima (A, B, C) and the corresponding protein conformations found along the transition have been depicted. The minimum A corresponds to the DFG-in state, in which the protein assumes a conformation where the Asp168 points towards the ATP binding site and the Phe169 is localized in an hydrophobic region. This minimum has a low value of z, indicating that here the protein assumes a conformation very similar to the crystallized one. On the contrary, the minimum C corresponds to a DFG-out state and it is at higher value of z. Here in order to better understand our results a comment is necessary. In fact, while the DFG-in state has been solved in p38 apo form, there is no structural evidence for the p38 apo DFG-out state.

Consequently we have used as final state of our path the X-ray structure of p38 in the DFG-out state complexed with an inhibitor. This can be tolerated since thanks to metadynamics and the use of PCV the system is able to explore conformations even far from the reference path by going at high values of z. This is what happened in our case with the minimum C that represents the most stable conformation of the p38 apo form in the DFG-out state. Minimum B represents the protein in an intermediate conformational state between the initial and final state. Here the DFG motif assumes the out state, while the rest of the activation loop is more similar to the conformation found in the DFG-in state. This conformation is typically found along the path when the system passes from the DFG-out to the DFG-in state. The stability of this pose has been assessed by means of tens ns of standard molecular dynamic simulation in which DFG motif does not exhibits any significant rearrangement. Interestingly, at this basin the activation loop and above all the binding site usually targeted by the ligands, have several elements of conformational diversity if compared to the initial and final states, thus representing a new potential target for future drug design. A particular comment has to be done for the broad minimum corresponding to $z = 8$. This large minimum is far away from the reference path and it does not correspond to one defined conformation of the protein. Here the activation loop assumes a large number of conformations which cannot be distinguished. This is due to its intrinsic fluctuations, confirmed by the numeous X-Ray structures in which this part of the protein is not resolved. Moreover also the NMR spectra rarely define this part of the protein[113]. The free energy profile as function of path collective variables shows the complexity of the process under investigation. A clearer representation of the DFG flip can be seen in Fig. 4.5 where the free energy profile has been reconstructed by means of the newly reweighting algorithm[125] as a function of the dihedral angles (ϕ and ψ) of the Asp residue. One can easily see that the DFG-in state corresponds to positive value of ϕ value, while the DFG-out state is localized at negative value of ϕ. The difference between the DFG-in/out states in free energy ΔF_{in-out} is approximately of -4 kcal/mol in favor of the DFG-in state. The similar difference in free energy of these two states is in line with the hypothesize of a conformational equilibrium between the DFG-in and DFG-out, as

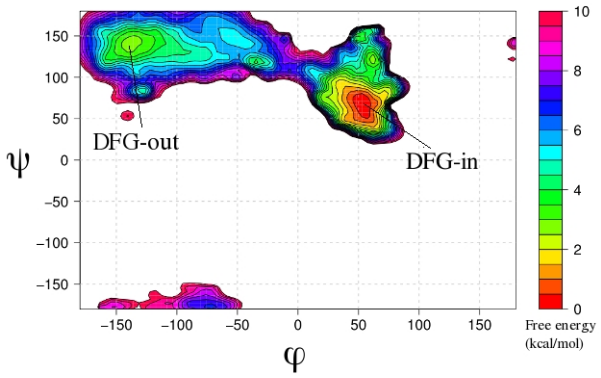

Figure 4.5: Free energy surface as a function of ϕ and ψ value of Asp for the Apo form of p38 kinase: the position of DFG-in and DFG-out is highlighted. Contour lines are plotted every 0.5 kcal/mol, color legend is in kcal/mol.

also suggested by the experiments.

4.4.2 DFG-in/out transition in ligand-bound form

The case of SB-203580

The DFG-flip has been studied also in the case of the protein in complex with SB-203580, a very well known p38 type I inhibitor. This ligand, as type I inhibitor, binds to the ATP binding site. It has three rings attached to a central imidazole: the pyridine ring N atom accepts a hydrogen bond from Met109 backbone in the hinge portion of the protein. The imidazole ring interacts with Lys53. The fluorophenyl ring occupies a hydrophobic pocket formed by a β-sheet of the N-terminal domain and αC-helix, it also interacts with the sidechains of Thr106 and Leu104, Leu75 and Leu86.

The 4-methylsulfinylphenyl moiety forms a stacking interaction with Tyr35. The methylsulfinyl group is near the activation loop, but makes no direct contacts with the protein.

The recent data about this ligand has made it an interesting case of ligand able to interact with both DFG-in and DFG-out states. The X-ray structure (2EWA) of the protein has

been resolved in both states. However, the whole activation loop beside the DFG motif is not solved, indicating once again a high flexibility of this part of the protein.

We used the thermalized structure of this complex in the DFG-in state as starting point for our metadynamics simulations. As reference path we used the same path adopted in the case of the apo protein. Along the transition mechanism, going from DFG-in to DFG-out state, the pyridine ring maintains its interaction with the Met109 and the fluorophenyl ring remains in its position, while the rest of the ligand fluctuates up and down as the protein goes from DFG-in to out state. The movement of the ligand follows the DFG-in/out transition, in fact as the Phe inserts into the ATP binding site, the ligand moves closer to the G-loop, interacting with Tyr35.

In fig. 4.6 the FES as function of s and z variables is shown. The DFG-in conformation

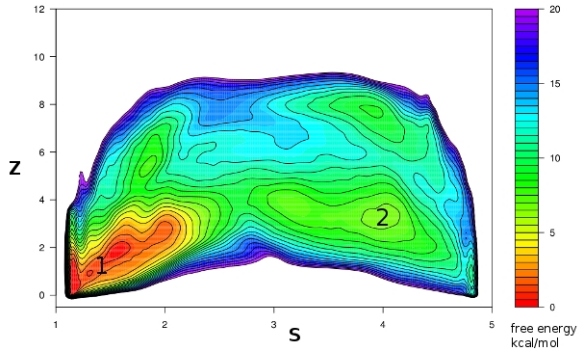

Figure 4.6: Free energy surface for P38 in complex with SB-203580 as a function of s and z variables. Contour lines are plotted every 1 kcal/mol, color legend is in kcal/mol.

corresponds to the first minimum, highlighted by number 1, while the DFG-out corresponds to minimum number 2. Given the high complexity of the FES as function of s and z variables, we report here the reconstructed FES as function of ϕ and ψ Asp168 dihedral angles. The reduction of the dimensionality of the FES representation is necessary because of the high intrinsic flexibility of the activation segment and it helps in highlighting the DFG transition in FES profile. Looking at Fig.4.7 we can easily see the existence of an

equilibrium between the two ending structures. In Fig. 4.8 the structure of the protein in the two main minima is reported.

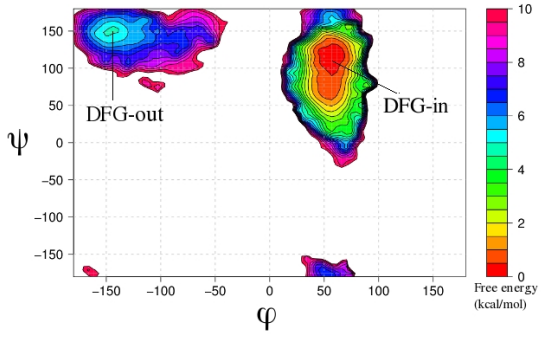

Figure 4.7: Free energy surface as a function of ϕ and ψ value of Asp in the case of protein complexed with SB-203580. Contour lines are plotted every 0.5 kcal/mol, color legend is in kcal/mol.

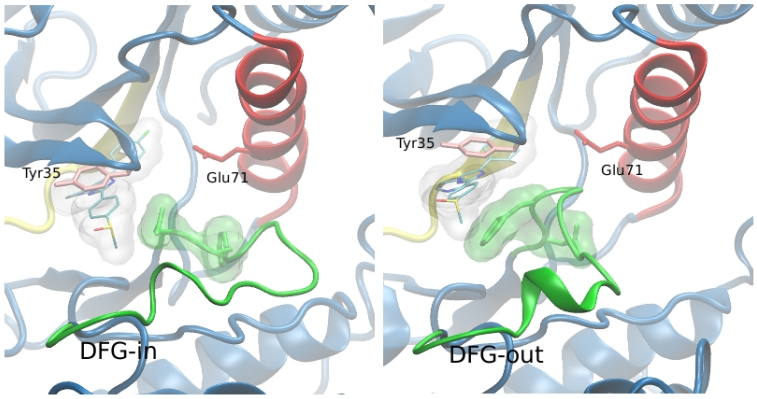

Figure 4.8: Representative structures of DFG-in and DFG-out conformations of the protein bound to the ligand found in the two main minima

Metadynamics has allowed the sampling of the whole process, and the found DFG-out state resembles that reported in the X-Ray structure. It is worthy nothing that the DFG-out state is stabilized by a π interaction between the phenyl ring of the ligand and aromatic residues such as Phe169 and Tyr35. Although experimental data shows the presence of an

equilibrium between DFG-in and DFG-out state, from our metadynamics calculation the existence of such equilibrium is confirmed, however it emerges that the DFG-in state is energetically more stable than the DFG-out conformation.

4.5 Conclusions

The use of enhanced sampling method has allowed us to describe at atomistic level the DFG-in to DFG-out transition in p38 kinase. This transition is important from a pharmaceutical point of view because this motion allows the creation of an additional binding pocket that can be targeted by many kinase inhibitors. In case of apo protein we have found an equilibrium between the two states. Experimental data have shown the presence of a similar equilibrium between DFG-in and DFG-out state also in the case of the protein bound to SB-203580, our metadynamics calculations confirm the existence of such equilibrium although it emerges that the DFG-in state is energetically more stable then the DFG-out conformation. The results obtained for this ligand make us optimistic about the possibility to study the transition also in the case of protein bound to BIRB-796. It will be interesting to understand the role of such ligand in this transition and to clarify the atomistic features of type II inhibitors along the DFG-in/out flip.

Chapter 5

Urea induced protein denaturation

5.1 Summary

In this chapter I present the study of folding/unfolding equilibrium for β-hairpin GB1 protein in both pure water and 8M urea solution. The chapter is organized as follows: firstly a brief introduction of the problem under investigation is given, then the technical details of the simulations are explained, finally the results are discussed. The use of parallel tempering together with metadynamics allowed us to characterize the nature of the unfolded state in the two solutions and to study the mechanism of action of urea.

5.2 Introduction

Among protein denaturants, urea is one of the most commonly used. In the last years a considerable progress has been made in understanding the molecular mechanism of urea induced protein denaturation through molecular dynamics simulations and experiments. Two mechanisms (see also the Introduction chapter) have been proposed based on a large number of experimental and theoretical studies. The indirect mechanism presumes that urea promotes ready solvation of hydrophobic groups by altering the water structure. The direct mechanism hypothesis suggests that urea unfolds protein through direct hydrogen bonding with the protein backbone, through electrostatic interactions with polar residues,

or through van der Waals attraction with the residues that sits at the protein surface. Different molecular dynamics studies have led to different views. The main reason for these different views arises from the fact that only partial unfolding events, starting from the native structures, have been reported even at elevated temperature. These limitations suggest that a more accurate sampling is needed to obtain a clear picture of the unfolding process and a deeper understanding of urea mechanism of action. Only recently, Canchi et al.[36] have used enhanced sampling technique in order to study the the folding/unfolding equilibrium of Trp-cage protein in the presence of urea. These authors used all-atom replica exchange molecular dynamics simulation and they captured the experimentally observed linear dependence of unfolding free energy on urea concentration.

Here we used the combination of metadynamics with parallel tempering in order to study at equilibrium the mechanism of action of urea in the folding/unfolding process of β-hairpin protein. Bussi and al.[42] have introduced for the first time the use of metadynamics together with parallel tempering technique. The use of these two methods allow to improve the capability of both methods: metadynamics improves the capability of parallel tempering to explore low probability regions, leading to a more reliable estimate of the height of the relevant free-energy barriers. On the other hand PT allows sampling of the degrees of freedom not explicitly included in the CVs, thus improving metadynamics accuracy. To understand in depth the nature of urea-protein interactions a newly developed reweighting algorithm that allows to recover the unbiased probability distribution of any variable from a well-tempered metadynamics simulation have been used[125].

We computed the free energy difference between the native and the unfolded states, both for the 8M solution and the pure water one. As expected, in urea solution the U$\rightleftharpoons$N equilibrium is shifted toward the unfolded state and it has more elongated conformations compared to what seen in pure water. For what concern the urea denaturation mechanism a preferential direct interaction between urea molecules and protein backbone has been observed. In particular, we have seen that at intermediate values of the radius of gyration there is a significant interaction through hydrogen bonds between backbone and urea as compared to water. While at higher value of radius of gyration the protein, urea and water

solvate all the other part of the protein.

5.3 Simulations details

5.3.1 Systems setup

Following the work of Bussi et al.[42], we used PTMETA to study the folding/unfolding equilibrium of β-hairpin protein in water and 8M urea. From now on we shall name water solution as WS and 8M urea one as US. The protein coordinates were obtained from the C-terminus (residues 41-56) of protein G (1G1B.pdb), the amino termini were capped: the N-terminal residue with acetil group (ACE) while the C-terminal residue with N-methyl amide capping group (NME). The protein was put in a rhombic dodecahedron box of ~ 84 nm^3 volume in both the systems: in the WS 1834 water molecules were added while in US urea system we used 266 urea molecules and 1196 water molecules. Three Na+ were also added to ensure charge neutrality in both systems. All simulations were performed using the GROMACS4 MD code[124] and the PLUMED[126] plugin with the Amber99SB force field[120] for the protein, the flexible version of OPLS-AA urea model for urea[127, 128] and TIP3P model for water. A preliminary minimization and NPT simulation for 2 ns at 300 K and 1 atm was performed on both systems. A time step of 2 fs was used. All covalent bonds were constrained to their equilibrium value using the Lincs algorithm[129]. The electrostatic interactions were calculated by the particle mesh Ewalds algorithm and the Lennard-Jones interactions with a cut-off of 0.9 nm was used.

5.3.2 PTMETA and the reweighting procedure

In the PTMETA technique multiple metadynamics simulations of the same system are performed in parallel. Each replica is simulated at different temperature and metadynamics bias is constructed on the same CVs. At fixed intervals, an exchange of configurations between two adjacent replicas is attempted on the basis of a Metropolis acceptance criterion. The exchange probability satisfies the detailed balance condition. By exchanging with higher temperatures, colder replicas are prevented from being trapped in local min-

ima. In this way the free-energy profile is filled in parallel at all temperatures.
Here we used 64 replicas of both systems in the temperature range 270-695 K. All replicas were simulated in NVT ensemble using a stochastic thermostat[130] with a coupling time of 0.1 ps. Exchanges were attempted every 0.1 ps. The resulting average acceptance probability was about 0.3 for all the replicas. Given the good results of Bussi et al.[42], we used the same CVs of their work. The first CV is the radius of gyration calculated on the hydrophobic core (R_{core}). This is composed by four aminoacids : Trp 43, Tyr 45, Phe 52 and Val 54. This collective variable is calculated using this formula:

$$R_{core} = \left(\frac{\sum_i^n |r_i - r_{com}|^2}{\sum_i^n m_i} \right)^{1/2} \tag{5.1}$$

where the sums are over the n atoms and the center of mass is defined by:

$$r_{com} = \frac{\sum_i^n r_i m_i}{\sum_i^n m_i} \tag{5.2}$$

Gyration radius is a common descriptor in protein folding studies because is able to discriminate between a completely unfolded protein and a molten globule state, where the protein is compact but disordered. In order to distinguish between molten globule and folded state we used as second CV the number of intra-molecular hydrogen bonds in the backbone. The number of backbone hydrogen bonds is evaluated using the switching function:

$$H_{back} = \sum_{ij} \frac{1 - \left(\frac{d_{ij}}{r_0}\right)^n}{1 - \left(\frac{d_{ij}}{r_0}\right)^m}. \tag{5.3}$$

Where r_0 is set to 2.5 Å, n and m to 6 and 12 respectively. i and j are the hydrogens and oxygens backbone atoms used to calculate the number of hydrogen bonds. We included in the CVs only the hydrogen bonds with a separation larger than 4 in the amino-acidic sequence in order to study the parallel β-sheet formations.
Well-tempered metadynamics parameters were set as follows: the bias factor was set at 10 for each replica, a Gaussian function was added every 1 ps with 0.05 and 0.025 width for the H_{back} and R_{core}. The initial Gaussian height was set to $1.0 * T/T_0$ kjoule/mol. Each replica has been simulated for 70 ns in case of WS and 100 ns of US.

In order to study the mechanism of urea denaturation, we have used the reweighting algorithm[125] that allows to recover the unbiased probability distribution of any variable from a well-tempered metadynamics simulation. In fact the introduction of a bias potential, if properly done, leads to the correct distribution for the chosen CVs, but distorts that of the other degrees of freedom. The method is able at reconstructing the Boltzmann distribution of a fast variable directly from the configurations produced in a well-tempered metadynamics run where a time-dependent bias potential is added on the slow degrees of freedom of the system. From this relation all quantities of interest can be evaluated. In particular quantitative comparison with experimental data is thus possible. In a previous work[131] the potential of this approach was shown by characterizing the conformational ensemble explored by a 13-residue helix-forming peptide by means of a well-tempered metadynamics/parallel tempering approach and comparing the reconstructed nuclear magnetic resonance scalar couplings with experimental data. The advantage of this reweighting procedure is not only to quantitative compare simulation with experimental data but also to save computer time because with this procedure any degrees of freedom different from CVs can be analyzed without performing additional metadynamics runs. In our case the use of reweighting algorithm has allowed us to estimate, by a post processing procedure, the FES for important variables related to the mechanism by which urea performs its action. For both WS and US we have studied the FES as a function of the following variables:

1. radius of gyration calculated on all the Cα (R_{ca}) and the distance from the crystallographic state in contact map space (D_{map}). For the R_{ca} we used the same formula as in Eq. 5.1 utilized to compute the radius of gyration for the hydrophobic core. The distance from the crystallographic state in contact map space was calculated as

$$D_{map} = \sqrt{\sum_{i}^{N_{cont}} (C_i(r) - C_i(Xray))^2} \tag{5.4}$$

where $C_i(r)$ is the ith contact for the configuration r and N_{cont} is the total number of contacts considered, and $C_i(Xray)$ is the ith contact for the crystallographic state. We considered all the contacts between the Cα atoms present in the crystallographic

structure and calculated as

$$\boldsymbol{C}(\boldsymbol{R}_{i,j}) = \frac{1-(\frac{r_{i,j}}{r_0})^p}{1-(\frac{r_{i,j}}{r_0})^q} \tag{5.5}$$

where d_{ij} is the distance between the ith and jth Cα atoms of the protein backbone, r_0 is taken to be 8.5 Å, n and m are set to 6 and 10 respectively.

2. the total number of hydrogen bonds between backbone and solvent ($HB_{bb-solv}$) and R_{ca}. The $HB_{bb-solv}$ has been calculated as the sum of contacts between backbone and water (HB_{bb-wat}) plus those between backbone and urea (HB_{bb-ure}) :

$$HB_{bb-solv} = HB_{bb-wat} + HB_{bb-ure} \tag{5.6}$$

In detail

$$HB_{bb-wat} = \sum_{Ob\_Hw} \frac{1-(d_{Ob\_Hw}/2.5)^{80}}{1-(d_{Ob\_Hw}/2.5)^{110}} + \sum_{Hb\_Ow} \frac{1-(d_{Hb\_Ow}/2.5)^{80}}{1-(d_{Hb\_Ow}/2.5)^{110}} \tag{5.7}$$

and

$$HB_{bb-ure} = \sum_{Ob\_Hu} \frac{1-(d_{Ob\_Hu}/2.5)^{80}}{1-(d_{Ob\_Hu}/2.5)^{110}} + \sum_{Hb\_Ou} \frac{1-(d_{Hb\_Ou}/2.5)^{80}}{1-(d_{Hb\_Ou}/2.5)^{110}} \tag{5.8}$$

where Ob and Hb are oxygens and hydrogens of the backbone, Ow and Hw oxygens and hydrogens of water and Ou and Hu oxygens and hydrogens of urea.

3. the number of all hydrogen bonds within the backbone (HB_{tot}) and R_{ca}. For HB_{tot} the same formula as in Eq. 5.3 has been used, but here all the possible donor-acceptor hydrogen bonds have been taken into account. The parameters were 2.5 Å for r_0, 20 and 80 for n and m.

4. the non native hydrogen bonds (HB_{nonnat}) calculated as the difference between HB_{tot} and the number of the native hydrogen bonds and R_{ca}. For the native hydrogen bonds we used the same formula as in Eq. 5.3 with r_0 = 2.5 Å, 20 and 80 for n and m. Only the six hydrogen bonds of the native structure have been considered in the sum.

5. the solvent accessible surface (SAS), calculated by means of g_sas tool of Gromacs[124] on the hydrophobic side chains of Trp, Tyr, Val and Phe.

6. in order to understand the urea mechanism of action we also reweighted in two separated FES the contribution of hydrogen bonds between backbone and water (HB_{bb-wat}) and those between backbone and urea (HB_{bb-ure}) as a function of R_{ca}.

5.4 Results and discussion

5.4.1 Free energy landscape

In Fig. 5.1 is reported the FES as a function of the radius of gyration calculated on the hydrophobic core (R_{core}) and the number of backbone hydrogen bonds (H_{back}) at room temperature, obtained by means of PTMETA for the two solutions under investigation. The FES was reconstructed from 100 ns of PTMETA simulation in the case of urea solution (US) and from 70 ns in the case of water solution (WS).
The shape of the FES in both cases is quite similar, and exhibits a L-shape. This indicates a zipping mechanism in which native contacts must first be broken before protein loses its hydrophobic core and goes into a stretched conformations. Going from the folded state (a) to the unfolded one (c), the protein breaks the hydrogen bonds of the backbone before loosing its hydrophobic core. A comparison of these FES and that reported by Bussi et al.[42] shows that the overall mechanism is quite similar, in fact also in that case there is a L-shape FES. However in our case, in both solutions, there is a minimum (b) that was not present in ref. [42]. We attribute this to the different force field used, Amber99SB [120] here and OPLS-AA[132] in ref.[42].

This minimum corresponds to structures intermediate between the native and the unfolded state: the protein has still some elements of secondary structures, in particular, the β-sheet hydrogen bonds close to the turn region are still formed. At values of H_{back}, smaller than 1, we can see in both cases the presence of a broad minimum that represents the unfolded state.

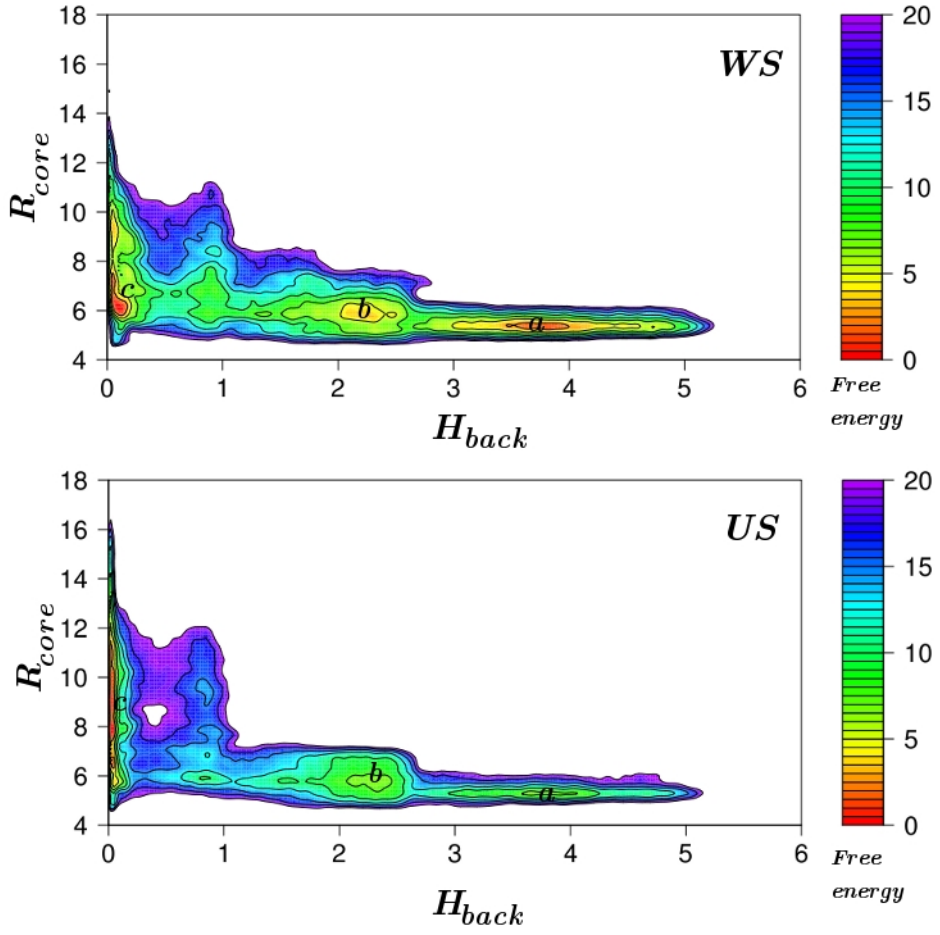

Figure 5.1: FES as a function of H_{back} of parallel β-sheet and R_{core} calculated on hydrophobic core of the protein: upper panel corresponds to WS, while lower panel to US. Contour lines are plotted every 5 kjoule/mol, color legend is in kjoule/mol.

The differences between the two solutions are to be seen in the depth of the minima. In fact, while the localization of the minima is almost similar in both cases, their depth is different. As expected, WS stabilizes the folded state, while US the unfolded one. From these FESs it is not trivial to understand the in the folding-unfolding equilibrium. In fact the CVs, while extremely useful at accelerating sampling, have a low resolution at small H_{back} values which is the region of the unfolded state. For these reasons we reconstructed[125] the FES as a function of different collective variables which are able to explore this region in greater details, namely the radius of gyration calculated on the Cα-atoms of the protein (R_{ca}) and the distance from the crystallographic state in contact map space (D_{map}) (see Simulations details). The resulting FESs are shown in Fig. 5.2 and in Fig. 5.3.

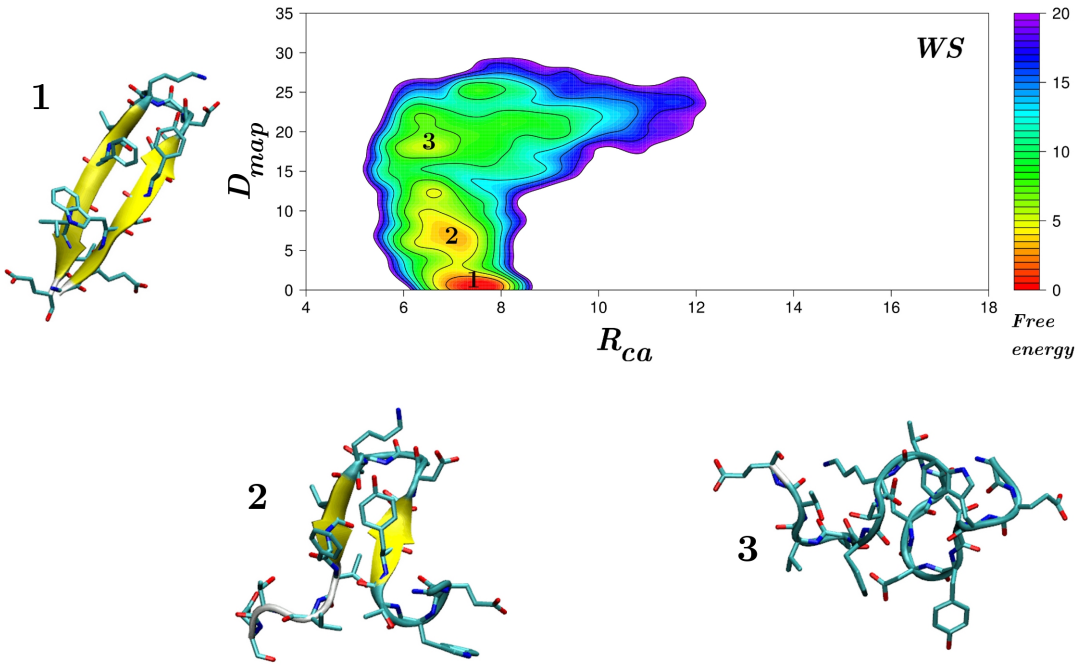

Figure 5.2: FES for WS as a function of distance from the crystallographic state in contact map space (D_{map}) and R_{ca} calculated on Cα atoms of the protein. Contour lines are plotted every 2 kjoule/mol, colour legend is in kjoule/mol. The subfigures 1 to 3 in WS show the typical structures encountered in such minima.

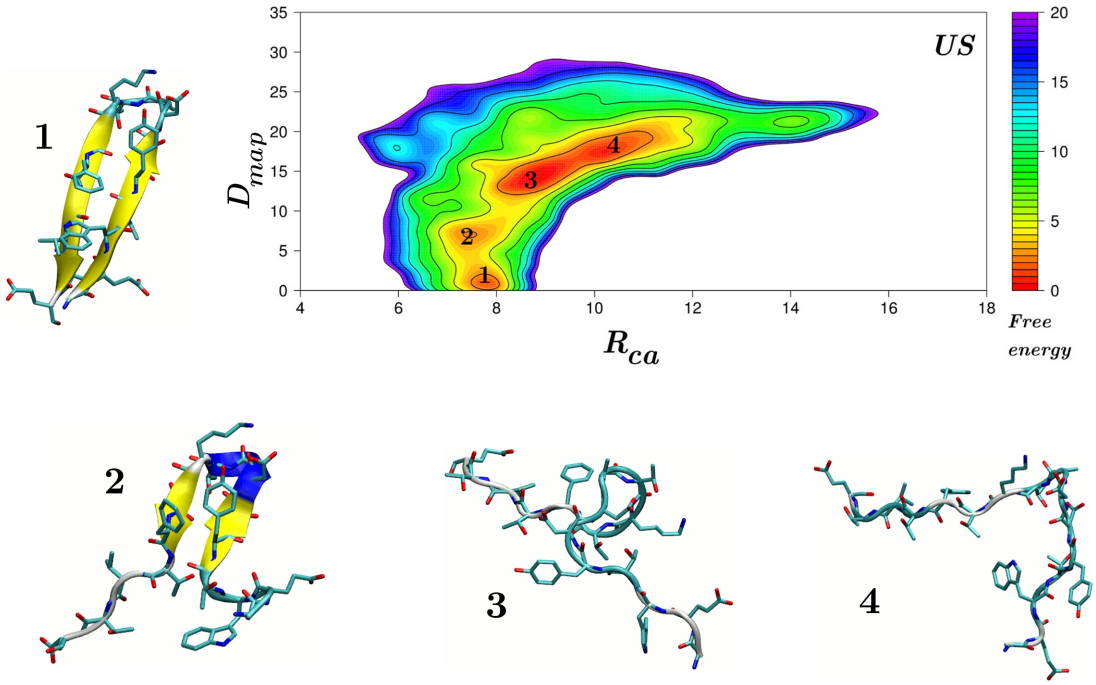

Figure 5.3: FES for US as a function of distance from crystallographic state in contact map space (D_{map}) and R_{ca} calculated on Cα atoms of the protein. Contour lines are plotted every 2 kjoule/mol, colour legend is in kjoule/mol. The subfigures 1 to 4 in US show the typical structures encountered in such minima.

It is to be seen in Fig. 5.2 that the basin a and b of Fig. 5.1 are preserved and are now mapped into minima 1 and 2. Basin c, as wished for, is now resolved in different local minima especially in US. As expected, the equilibrium between folded and unfolded states is shifted in US towards the disordered state. According to this FESs, the folding free energy ΔF_{FU} is $\sim$0 in WS and $\sim$1 kcal/mol in US. These results are in remarkable agreement with NMR experiments which suggest that the population of β-hairpin structure is 42 % in water [133] and 32 % in 6 M urea[134]. The ensemble of unfolded state is different in the two cases and it is seen that the gyration radius is larger in urea and several likely structures can be found with radius of gyrations that range from 8 to 12 Å. Several experiments have shown that unfolded state in urea solution is considerably more extended than when de-

naturation is induced by other agents such as pH or temperature[135, 136, 137]. Moreover recent single-molecule FRET experiments showed an increase of the radius of gyration of unfolded state for different proteins with increasing denaturant concentration[138]. This qualitative agreement between theory and experiment reassures us on the validity of our model. Thus with some confidence we can proceed further in the analysis in order to get a deeper insight in the role of urea.

5.4.2 The role of the solvent

An important insight into the nature of the unfolding mechanism comes from the study of the interaction between solvent and protein. Firstly we examine the interactions between protein backbone and solvent in terms of number of hydrogen bonds, then we report the effect of the solvent on the packing of the hydrophobic side chain. In Fig. 5.4-A and B we report the FES as a function of the number of hydrogen bonds between the backbone and the solvent ($HB_{bb-solv}$), the distance from the crystallographic structure in contact map space (D_{map}) and Cα gyration radius (R_{ca}). We consider all the possible hydrogen bonds between backbone and solvent in both solutions (see Simulation details). The relevance of these FES is due to the number of contacts between the protein backbone and the solvent in the two solutions studied. From the panel A and B it is clear that there is a different pattern of solvation and that the backbone polar group in the unfolded state are more solvated in US. We further analyzed the solution as a function of the gyration radius (Fig. 5.4-C). In US even the compact states are more solvated and this trend is further enhanced at intermediate values of R_{ca} where the protein is partially open. When the protein is totally stretched this difference tends to be much reduced. It is also instructive to analyze the total number of intramolecular hydrogen bonds and that of non native contacts (Fig.5.4-D). In particular in water the number of internal hydrogen bonds decays more slowly than in US and in the intermediate R_{ca} regime a larger number of nonnative contacts is found. This indicates a greater resilience of the water system to solvation and suggests the existence of different intermediate structures along the folding pathways in the two systems. We also studied the effect of the solvent on the hydrophobic core in

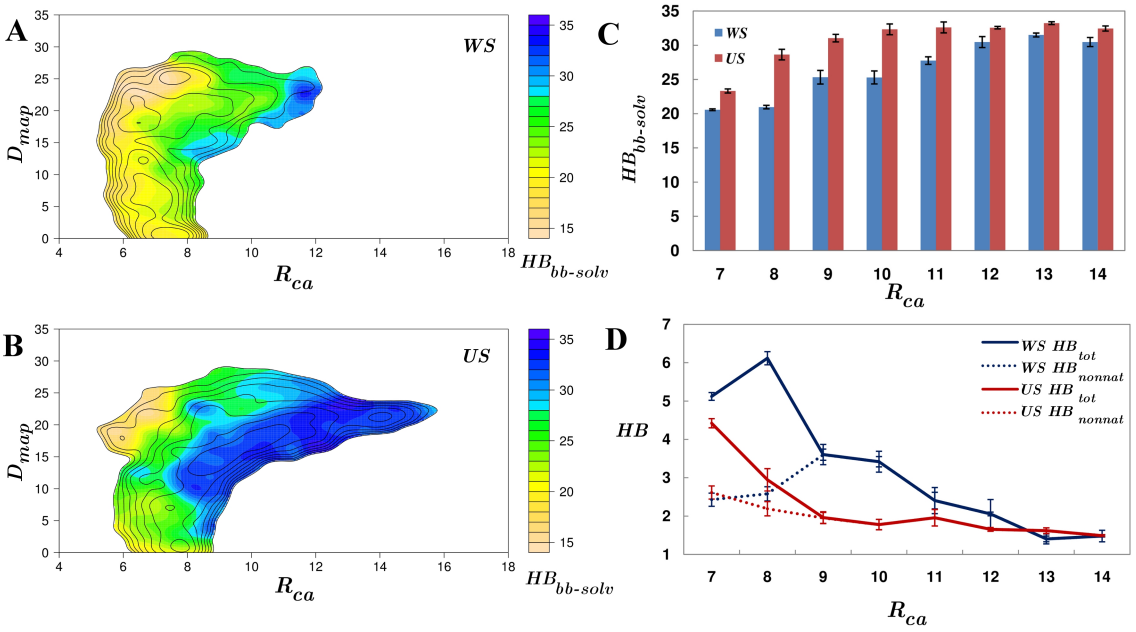

Figure 5.4: The role of the solvent: comparison of the FES as a function of the number of hydrogen bonds between backbone and solvent, R_{ca} and D_{map} in WS (A) and US (B). Contour lines are plotted every 2 kjoule/mol, color legend is in kjoule/mol. The histogram in C represents the averaged number of hydrogen bonds in the two solutions, blue corresponds to WS and red to US. In D is reported the averaged number of hydrogen bonds within the backbone: blue states for WS and red for US. In solid lines the total number of hydrogen bonds are reported, while in dashed line the non native hydrogen bonds within the backbone can be seen.

the two solutions. The original FES as function of the solvent accessible surfaces of the hydrophobic core (SAS), the distance from the crystallographic state in contact map space (D_{map}) and the R_{ca} have been reweighted. The SAS of hydrophobic core describes how the hydrophobic residues are coordinated in the space. Low values of SAS correspond to highly packed hydrophobic core, while high values mean the loss of the core.

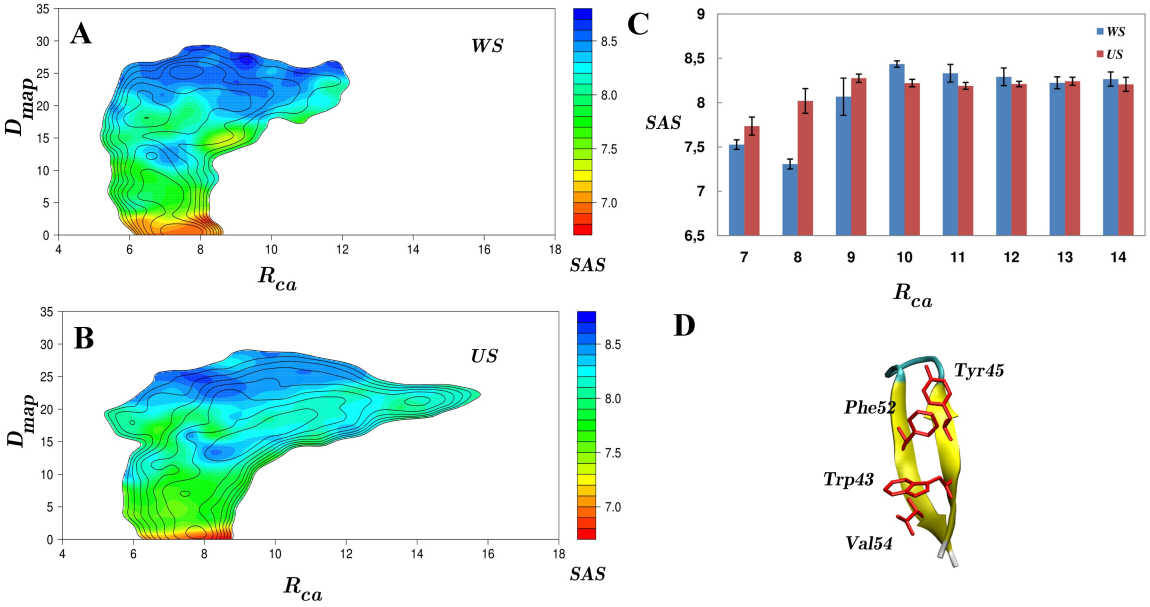

Figure 5.5: A and B depict the FES as a function of the solvent accessible surface of the hydrophobic core, R_{ca} and D_{map} for WS and US. Contour lines are plotted every 2 kjoule/mol, colour legend is in kjoule/mol. Panel C shows the averaged coordination of hydrophobic core along the progress of R_{ca}. In D the native state of β-hairpin protein is reported, in licorice the four aminoacids that form the hydrophobic core are highlighted.

When the protein backbone is close to its crystallographic structure (low value of D_{map}) the average conformation of the hydrophobic core is different in the two solutions. In particular, it can be seen that in WS the core is preferentially in a closed conformation. Whereas in US the presence of urea prevents the collapse of the core and stabilizes an unpaired conformation of the hydrophobic side chains. In other region of the conformational space the behavior of the hydrophobic core is rather similar in the two solutions. This trend

is confirmed by the average value of SAS as a function of R_{ca} (see Fig.5.5-C). This is not surprising because of the nature of the hydrophobic core and the small size of the protein. In this respect the behavior of the two systems appear in first approximation to be similar. As already mentioned in the Introduction, in literature there is an open issue about which is the main driving force in urea denaturation. While many computational studies provide increasing evidence for direct interactions as the primary driving force for denaturation, it is discussed whether either polar or apolar interactions between urea and protein dominate. In particular the relative importance of protein-solvent H-bonds and Van der Waals interactions in the unfolding process has been discussed [29, 27, 30, 31, 32, 33, 34]. Our data clearly indicate that the denaturation of β-hairpin by urea mainly involves the direct interaction between the solvent and the backbone polar groups. While the formation of hydrogen bonds between the protein and the solvent seems to be the driving force of unfolding mechanism in urea, a smaller role can be assigned to hydrophobic effect.

5.4.3 Urea's mechanism of action

In order to evaluate which is the main force that enables and stabilizes the formation of such unfolded state in US, we have studied the role of the solvent within US: we report in Fig. 5.6 the total number of hydrogen bond per species. In analyzing these data one must keep in mind that while urea can form a larger number of hydrogen bonds, its concentration is much smaller. Thus a comparable number of hydrogen bonds indicates a stronger propensity of urea molecules to bind to the protein backbone. In many studies it has been demonstrated experimentally that urea participates in hydrogen bonds with peptide group [28].

Going from the folded to the unfolded state the whole range of R_{ca} can be split into three main regions. The first one, A, corresponds to very compact conformations. Within this region lies also the folded state. The interactions between urea and water molecules are comparable. This indicates a similar propensity of the backbone to form hydrogen bonds with both urea and water molecules. Going into the second region, where R_{ca} goes from 8 to 9 Å there is small preponderance of the hydrogen bonds with urea. While in

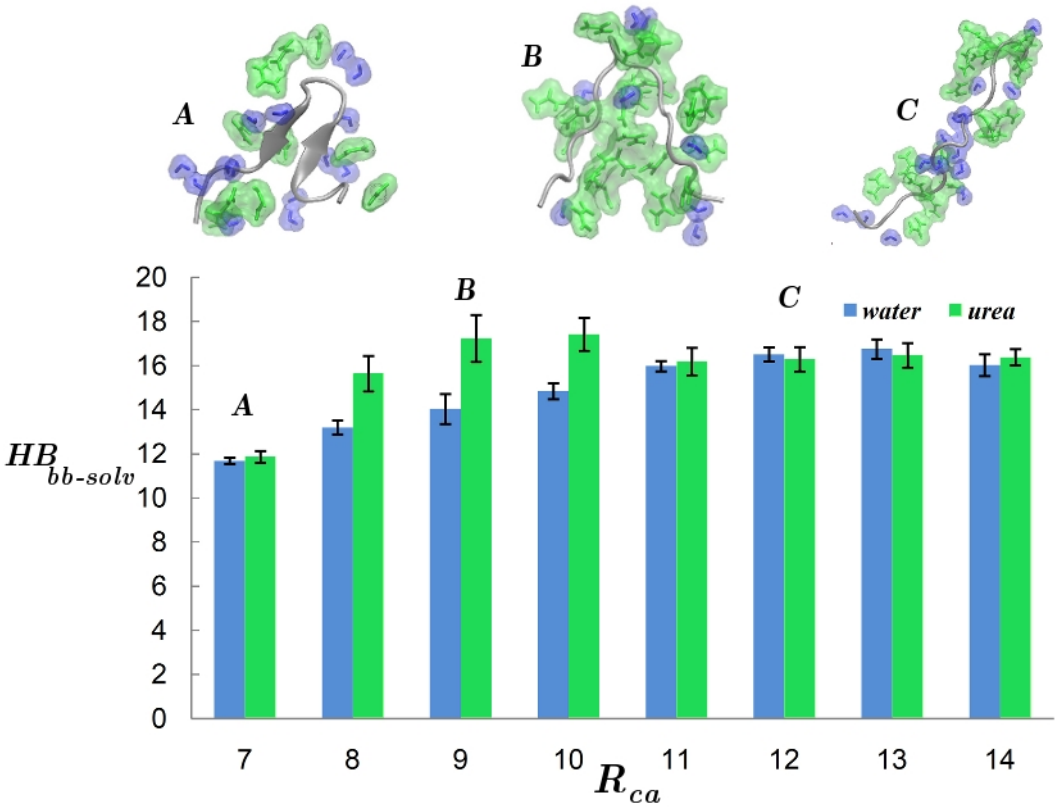

Figure 5.6: Averaged number of hydrogen bonds formed between backbone and urea is reported in green, in blue those between backbone and water. Three typical structures encountered at value of 7, 9 and 12 of R_{ca} are reported in figure A, B, and C

the third region, C, the interactions of water and urea with backbone are again similar. Representative snapshots of the environment behavior around the protein at value of 7, 9, 12 Å of R_{ca} can be seen in the upper panel of 5.6. This behavior, along the unfolding pathway, means that when the protein starts to unfold and reaches a more elongated conformation, there is a paucity of water molecules around it, while urea can easily interact. At higher value of R_{ca} there is no more this pronounced preferential solvation of the protein by urea. The existence of such preferential interaction with urea along the unfolding pathway has been already seen in other studies[35, 139]. In particular a very long molecular dynamics simulation of lysozyme in 8M urea has shown the formation of a dry globule state in which there is a paucity of water around the protein. Here, equilibrium simulation of the

folding/unfolding pathway of β-hairpin suggests the presence of a state at intermediate value of R_{ca} in between the folded and completely stretched state in which the protein preferentially interacts with urea through backbone hydrogen bonds.

5.5 Conclusions

Here, the use of advanced sampling techniques allowed us a deeper understanding of the main differences in the reversible folding-unfolding free-energy landscape of β-hairpin GB1 in 8 M urea solution and pure water. The folding free energy ΔF_{FU} for the two solutions has been calculated, confirming NMR experiments: as expected, the equilibrium between folded and unfolded states is shifted in US towards the disordered one. The nature of the unfolded states has been extensively analyzed: our data show that urea stabilizes the protein in a completely unfolded conformation, with high value of gyration radius. This behavior confirms that the addition of denaturants allows the protein to unfold in a stretched conformation. In the absence of denaturant the unfolded state is in equilibrium with native state and it is noticeably more compact. For what concern the mechanism by which urea acts as denaturant, a preferential direct interaction between urea molecules and protein backbone has been found. At intermediate value of R_{ca} along the unfolded pathway we have found a preferential interaction of protein backbone with urea molecules in agreement with previous MD simulations. At higher value of gyration urea and water solvate all the other part of the protein, allowing it to unfold in a completely stretched conformation.

Chapter 6

Conclusion

Studying rare events such as protein folding or protein conformational plasticity with MD poses several challenges which are related to the difference between the time scale accessible by MD simulation and that in which interesting events take place.

During my PhD I have applied metadynamics to study different rare events related to biomolecular flexibility. Metadynamics is able both to accelerate the sampling and to reconstruct free-energy profiles along selected degrees of freedom at the same time.

In particular the use of metadynamics has highlighted some of the dynamical features of kinase conformational plasticity. In these proteins, conformational motion, which allows reversible switching between distinct states, is a central feature of their regulatory mechanisms. While our current understanding of kinase regulation owes much to experimental informations coming from X-ray crystallography, static pictures must be combined with information from other approaches that more effectively capture the dynamics of the switching process.

In Chapters 3 and 4 the atomistic details associated with the inactive mechanism of CDK5 and p38α kinase have been presented. These proteins are representative of two distinct kinase inactivation mechanisms: CDK5 can access inactive conformations through a transition between αC-helix Glu-in and αC-helix Glu-out conformations, while the flip of the well conserved DFG motif allows p38α kinase to access alternative conformations.

Metadynamics combined with the path collective variable method has provided details of the open-to-closed CDK5 conformational switch that involves large scale motions of the αC-helix and the T-loop. A two step mechanism along the transition has been shown: first, the αC-helix rotates by $\sim$45°, allowing the interaction between Glu51 and Arg149; then the CDK5 activation loop refolds to assume the closed conformation. Along the open-to-closed path, one meta-stable conformation that shows unique structural features has been identified.

In the case of p38α kinase metadynamics has revealed the main features of the motion from DFG-in to DFG-out state. The equilibrium between the two states has been studied also in the case of protein bound to an inhibitor, SB-203580. The presence of SB-203580, a type I kinase inhibitor, does not interfere too much with the DFG-in/out flip of this protein as expected from experimental data. From these results the dynamical aspects of the inactivation mechanism of these two kinases have been disclosed. The data obtained here will be of precious help for the design of potent and selective kinase inhibitors.

Finally, in Chapter 5 the use of the combined technique such as metadynamics together with parallel tempering has allowed us to study the equilibrium properties of the folding-unfolding mechanism of β-hairpin in pure water and 8M urea solution. While standard MD simulation could in principle be used to characterize the early events of protein folding, the use of methods such as PTMETA to speed-up the dynamics of the system under investigation is very useful to obtain more informations about the mechanism of folding and the role of denaturants such as urea. The nature of the unfolded state in different environments has been extensively analyzed: in particular, our data show that urea stabilizes the protein in a completely unfolded conformation, with high value of gyration radius. In the pure water solution the unfolded state is in equilibrium with native state and it is noticeably more compact than the urea-unfolded one. We found that denaturation occurs predominantly by direct interaction of urea with the protein backbone through hydrogen bonds. At intermediate value of radius of gyration the protein preferentially interacts with urea as compared to water. While at higher value of radius of gyration, urea and water solvate all the other part of the protein, allowing it to unfold in a completely stretched

conformation.

Acknowledgements

First of all, I am particularly grateful to my supervisor, Professor Michele Parrinello, for giving me the opportunity to work in his group. I have greatly appreciated his advice, his suggestions and his support in these years.
I am grateful to Francesco Gervasio and Davide Branduardi for the nice collaboration on the investigation of kinase conformational plasticity. I have learnt a lot from both of you.
Special thanks to Vittorio Limongelli for the great atmosphere we have created during the study of p38 inactivation mechanism. Thanks not only for our work related discussions about science but also about many other topics more related to everyday life.
I am also indebted to Alessandro Barducci for his support, encouragement and good ideas about our work on urea unfolding protein. It has been a good time spent together on this subject!
I would also like to thank all the current and former members of the group for the stimulating discussions and the support during these years, and for the nice after-hours events.
A particular thank to Giacomo Miceli for having shared with me our worries for the future, doubts, bad moments during these last years.
Daniela Wirz also deserves special thanks for her understanding, support and encouragement, and for creating and preserving a nice working atmosphere in the group.
Finally Davide deserves my biggest thanks and gratitude for his huge support and for his confidence in me.

List of Figures

Abbreviations and symbols

β	Thermodynamic beta, $\beta = 1/(k_B T)$
β-hairpin	16-residue C-terminal fragment of protein GB1
$\boldsymbol{R}$	Coordinates of the system, $\boldsymbol{R} = (\boldsymbol{R}_1, ..., \boldsymbol{R}_N)$
$\boldsymbol{S}$	Set of collective variables, $\boldsymbol{S}(\boldsymbol{R}) = (S_1(\boldsymbol{R}), ..., S_d(\boldsymbol{R}))$
ω	Metadynamics energy rate
σ_i	Gaussian width for the ith CV
τ_G	Gaussian deposition stride
k_B	Boltzmann constant
$s(\boldsymbol{R})$	s variable of PCV
US	8M Urea solution
V_G	Metadynamics bias potential, $V_G = V_G(\boldsymbol{S}, t)$
W	Gaussian height
WS	Water solution
$z(\boldsymbol{R})$	z variable of PCV
C-lobe	C-terminal lobe

CDK2	Cyclin Dependent Kinase 2
CDK5	Cyclin Dependent Kinase 5
CMAP	Contact Map
CV	Collective Variable
DFG	Aspartate-Phenilalanine-Glycine
FES	Free-energy surface
FRET	Fluorescence Resonance Energy Transfer
MD	Molecular Dynamics
N	Native state
N-lobe	N-terminal lobe
NPT	Ensemble at constant number of particles (N), pressure (V) and temperature (T)
NVT	Ensemble at constant number of particles (N), volume (V) and temperature (T) or canonical ensemble
p38	Mitogen-activated protein kinase p38 α
PCV	Path Collective Variables
PT	Parallel Tempering
PTMETA	Parallel Tempering and metadynamics combined
U	Unfolded state

Bibliography

[1] K. Henzler-Wildman and D. Kern. Dynamic personalities of proteins. *Nature*, 450(7172):964–972, 2007.

[2] M. Karplus and J.A. McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural & Molecular Biology*, 9(9):646–652, 2002.

[3] A. Laio and M. Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12562, 2002.

[4] A. Barducci, G. Bussi, and M. Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical review letters*, 100(2):20603, 2008.

[5] G. Manning, D.B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. The protein kinase complement of the human genome. *Science*, 298(5600):1912, 2002.

[6] P. Cohen. Protein kinases: the major drug targets of the twenty-first century? *Nature Reviews Drug Discovery*, 1(4):309–315, 2002.

[7] M. Rabiller, M. Getlik, S. Klüter, A. Richters, S. Tückmantel, J.R. Simard, and D. Rauh. Proteus in the World of Proteins: Conformational Changes in Protein Kinases. *Archiv der Pharmazie*, 343(4):193–206, 2010.

[8] L.N. Johnson. Protein kinase inhibitors: contributions from structure to clinical compounds. *Quarterly Reviews of Biophysics*, 42(1):1–40, 2009.

[9] L.N. Johnson, M.E.M. Noble, and D.J. Owen. Active and inactive protein kinases: structural basis for regulation. *Cell*, 85(2):149–158, 1996.

[10] M. Huse and J. Kuriyan. The conformational plasticity of protein kinases. *Cell*, 109(3):275–282, 2002.

[11] E.R. Wood, A.T. Truesdale, O.B. McDonald, D. Yuan, A. Hassell, S.H. Dickerson, B. Ellis, C. Pennisi, E. Horne, K. Lackey, et al. A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib). *Cancer research*, 64(18):6652, 2004.

[12] B. Nagar, W.G. Bornmann, P. Pellicena, T. Schindler, D.R. Veach, W.T. Miller, B. Clarkson, and J. Kuriyan. Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer research*, 62(15):4236, 2002.

[13] S. Wilhelm, C. Carter, M. Lynch, T. Lowinger, J. Dumas, R.A. Smith, B. Schwartz, R. Simantov, and S. Kelley. Discovery and development of sorafenib: a multikinase inhibitor for treating cancer. *Nature Reviews Drug Discovery*, 5(10):835–844, 2006.

[14] A.A. Russo, L. Tong, J.O. Lee, P.D. Jeffrey, and N.P. Pavletich. Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature*, 395(6699):237–243, 1998.

[15] N.M. Levinson, O. Kuchment, K. Shen, M.A. Young, M. Koldobskiy, M. Karplus, P.A. Cole, and J. Kuriyan. A Src-like inactive conformation in the abl tyrosine kinase domain. *PLoS Biology*, 4(5):753, 2006.

[16] J.H. Weishaupt, C. Neusch, and M. Bähr. Cyclin-dependent kinase 5 (CDK5) and neuronal cell death. *Cell and tissue research*, 312(1):1–8, 2003.

[17] A. Berteotti, A. Cavalli, D. Branduardi, F.L. Gervasio, M. Recanatini, and M. Parrinello. Protein conformational transitions: the closure mechanism of a kinase

explored by atomistic simulations. *Journal of the American Chemical Society*, 131(1):244–250, 2009.

[18] R. Diskin, D. Engelberg, and O. Livnah. A novel lipid binding site formed by the MAP Kinase insert in p38α. *Journal of molecular biology*, 375(1):70–79, 2008.

[19] D.W. Bolen and G.D. Rose. Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annual review of biochemistry*, 77:339–362, 2008.

[20] J.A. Rupley. The effect of urea and amides upon water structure. *Journal of Physical Chemistry*, 68(7):2002–2003, 1964.

[21] H.S. Frank and F. Franks. Structural approach to the solvent power of water for hydrocarbons; urea as a structure breaker. *Journal of Chemical Physics*, 48:4746, 1968.

[22] F. Vanzi, B. Madan, and K. Sharp. Effect of the protein denaturants urea and guanidinium on water structure: A structural and thermodynamic study. *Journal of the American Chemical Society*, 120(41):10748–10753, 1998.

[23] B.J. Bennion and V. Daggett. The molecular basis for the chemical denaturation of proteins by urea. *Proceedings of the National Academy of Sciences of the United States of America*, 100(9):5142, 2003.

[24] A. Caballero-Herrera, K. Nordstrand, K.D. Berndt, and L. Nilsson. Effect of urea on peptide conformation in water: molecular dynamics and experimental characterization. *Biophysical journal*, 89(2):842–857, 2005.

[25] G.I. Makhatadze and P.L. Privalov. Protein interactions with urea and guanidinium chloride: A calorimetric study. *Journal of molecular biology*, 226(2):491–505, 1992.

[26] R.D. Mountain and D. Thirumalai. Molecular dynamics simulations of end-to-end contact formation in hydrocarbon chains in water and aqueous urea solution. *Journal of the American Chemical Society*, 125(7):1950–1957, 2003.

[27] E.P. O'Brien, R.I. Dima, B. Brooks, and D. Thirumalai. Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: lessons for protein denaturation mechanism. *Journal of the American Chemical Society*, 129(23):7346–7353, 2007.

[28] W.K. Lim, J. Rösgen, and S.W. Englander. Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proceedings of the National Academy of Sciences*, 106(8):2595, 2009.

[29] A. Wallqvist, DG Covell, and D. Thirumalai. Hydrophobic interactions in aqueous urea solutions with implications for the mechanism of protein denaturation. *Journal of the American Chemical Society*, 120(2):427–428, 1998.

[30] J.L. England, D. Lucent, and V.S. Pande. A role for confined water in chaperonin function. *Journal of the American Chemical Society*, 130(36):11838–11839, 2008.

[31] R. Zangi, R. Zhou, and BJ Berne. Urea's action on hydrophobic interactions. *Journal of the American Chemical Society*, 131(4):1535–1541, 2009.

[32] M.C. Stumpe and H. Grubmüller. Interaction of urea with amino acids: implications for urea-induced protein denaturation. *Journal of the American Chemical Society*, 129(51):16126–16131, 2007.

[33] M.C. Stumpe and H. Grubmüller. Polar or apolar-the role of polarity for urea-induced protein denaturation. *PLoS Computational Biology*, 4(11):e1000221, 2008.

[34] M.C. Stumpe and H. Grubmüller. Urea impedes the hydrophobic collapse of partially unfolded proteins. *Biophysical journal*, 96(9):3744–3752, 2009.

[35] L. Hua, R. Zhou, D. Thirumalai, and B.J. Berne. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proceedings of the National Academy of Sciences*, 105(44):16928, 2008.

[36] D.R. Canchi, D. Paschek, and A.E. Garcia. Equilibrium Study of Protein Denaturation by Urea. *Journal of the American Chemical Society*, 132(7):2338–2344, 2010.

[37] J.A. McCammon, B.R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, 1977.

[38] A. Laio and F.L. Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics*, 71:126601, 2008.

[39] D. Branduardi, F.L. Gervasio, and M. Parrinello. From A to B in free energy space. *The Journal of chemical physics*, 126:054103, 2007.

[40] M. Bonomi, D. Branduardi, F.L. Gervasio, and M. Parrinello. The unfolded ensemble and folding mechanism of the C-terminal GB1 β-hairpin. *Journal of the American Chemical Society*, 130(42):13938–13944, 2008.

[41] M. Vendruscolo, R. Najmanovich, and E. Domany. Protein folding in contact map space. *Physical Review Letters*, 82(3):656–659, 1999.

[42] G. Bussi, F.L. Gervasio, A. Laio, and M. Parrinello. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *Journal of the American Chemical Society*, 128(41):13435–13441, 2006.

[43] D.O. Morgan and H.L. De Bondt. Protein kinase regulation: insights from crystal structure analysis. *Current opinion in cell biology*, 6(2):239–246, 1994.

[44] J.C. Williams and A.E. McDermott. Dynamics of the Flexible Loop of Triose-Phosphate Isomerase: The Loop Motion Is Not Ligand Gated. *Biochemistry*, 34(26):8309–8319, 1995.

[45] D.D. Boehr, D. McElheny, H.J. Dyson, and P.E. Wright. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science*, 313(5793):1638, 2006.

[46] A.G. Palmer III. NMR characterization of the dynamics of biomacromolecules. *Chemical reviews*, 104(8):3623–3640, 2004.

[47] Q. Cui and M. Karplus. Catalysis and specificity in enzymes: a study of triosephosphate isomerase and comparison with methyl glyoxal synthase. *Advances in Protein Chemistry*, 66:315–372, 2003.

[48] E.Z. Eisenmesser, O. Millet, W. Labeikovsky, D.M. Korzhnev, M. Wolf-Watz, D.A. Bosco, J.J. Skalicky, L.E. Kay, and D. Kern. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 438(7064):117–121, 2005.

[49] M. Wolf-Watz, V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E.Z. Eisenmesser, and D. Kern. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nature structural & molecular biology*, 11(10):945–949, 2004.

[50] S.C. Blanchard, R.L. Gonzalez, H.D. Kim, S. Chu, and J.D. Puglisi. tRNA selection and kinetic proofreading in translation. *Nature structural & molecular biology*, 11(10):1008–1014, 2004.

[51] T. Ha. Single-molecule fluorescence resonance energy transfer. *Methods*, 25(1):78–86, 2001.

[52] S. Myong, B.C. Stevens, and T. Ha. Bridging conformational dynamics and function using single-molecule spectroscopy. *Structure*, 14(4):633–643, 2006.

[53] P.J. Rothwell, S. Berger, O. Kensch, S. Felekyan, M. Antonik, B.M. Wöhrl, T. Restle, R.S. Goody, and C.A.M. Seidel. Multiparameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase: primer/template complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 100(4):1655, 2003.

[54] B. Schuler, E.A. Lipman, and W.A. Eaton. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, 419(6908):743–747, 2002.

[55] Z. Zhang, P.T. Rajagopalan, T. Selzer, S.J. Benkovic, and G.G. Hammes. Single-molecule and transient kinetics investigation of the interaction of dihydrofolate re-

ductase with NADPH and dihydrofolate. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2764, 2004.

[56] J.A. McCammon, B.R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, 1977.

[57] R. Friedman. Proton Transfer on the Molecular Surface of Proteins and Model Systems. *Israel Journal of Chemistry*, 49(2):149–153, 2009.

[58] C.P. Barrett and M.E.M. Noble. Molecular motions of human cyclin-dependent kinase 2. *Journal of Biological Chemistry*, 280(14):13993, 2005.

[59] V. Tozzini. Coarse-grained models for proteins. *Current opinion in structural biology*, 15(2):144–150, 2005.

[60] S. Yang and B. Roux. Src kinase conformational activation: Thermodynamics, pathways, and mechanisms. *PLoS Computational Biology*, 4:e1000047, 2008.

[61] P.C. Whitford, O. Miyashita, Y. Levy, and J.N. Onuchic. Conformational transitions of adenylate kinase: switching by cracking. *Journal of molecular biology*, 366(5):1661–1671, 2007.

[62] Q. Lu and J. Wang. Single molecule conformational dynamics of adenylate kinase: energy landscape, structural correlations, and transition state ensembles. *Journal of the American Chemical Society*, 130(14):4772–4783, 2008.

[63] D.O. Morgan. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annual review of cell and developmental biology*, 13(1):261–291, 1997.

[64] Y. Lee and P. Sicinski. Targeting cyclins and cyclin-dependent kinases in cancer: Lessons from mice and hopes for therapeutic applications in human. *Cell cycle*, 5(18):2110–2114, 2006.

[65] J.A. Geyer, S.T. Prigge, and N.C. Waters. Targeting malaria with specific CDK inhibitors. *Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics*, 1754(1-2):160–170, 2005.

[66] L.M. Schang. Advances on cyclin-dependent kinases (CDKs) as novel targets for antiviral drugs. *Current Drug Targets-Infectious Disorders*, 5(1):29–37, 2005.

[67] EA Monaco 3rd and M.L. Vallano. Role of protein kinases in neurodegenerative disease: cyclin-dependent kinases in Alzheimer's disease. *Frontiers in Bioscience*, 10:143–159, 2005.

[68] N.P. Pavletich. Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. *Journal of molecular biology*, 287:821–828, 1999.

[69] A.J. Obaya and J.M. Sedivy. Regulation of cyclin-Cdk activity in mammalian cells. *Cellular and Molecular Life Sciences*, 59(1):126–142, 2002.

[70] K. Ishiguro, S. Kobayashi, A. Onion, M. Takamatsu, S. Yonekura, K. Anzai, K. Imahori, and T. Uchida. Identification of the 23 kDa subunit of tau protein kinase II as a putative activator of cdk5 in bovine brain. *FEBS letters*, 342(2):203–208, 1994.

[71] R. Dhavan and L.H. Tsai. A decade of CDK5. *Nature Reviews Molecular Cell Biology*, 2(10):749–759, 2001.

[72] Z.H. Cheung, A.K.Y. Fu, and N.Y. Ip. Synaptic roles of Cdk5: implications in higher cognitive functions and neurodegenerative diseases. *Neuron*, 50(1):13–18, 2006.

[73] J.C. Cruz and L.H. Tsai. Cdk5 deregulation in the pathogenesis of Alzheimer's disease. *Trends in Molecular Medicine*, 10(9):452–458, 2004.

[74] P.D. Smith, S.J. Crocker, V. Jackson-Lewis, K.L. Jordan-Sciutto, S. Hayley, M.P. Mount, M.J. O'Hare, S. Callaghan, R.S. Slack, S. Przedborski, et al. Cyclin-dependent kinase 5 is a mediator of dopaminergic neuron loss in a mouse model of

Parkinson's disease. *Proceedings of the National Academy of Sciences of the United States of America*, 100(23):13650, 2003.

[75] M.D. Nguyen and J.P. Julien. Cyclin-dependent kinase 5 in amyotrophic lateral sclerosis. *Neurosignals*, 12(4-5):215–220, 2003.

[76] J.L. Hallows, R.E. Iosif, R.D. Biasell, and I. Vincent. p35/p25 is not essential for tau and cytoskeletal pathology or neuronal loss in Niemann-Pick type C disease. *Journal of Neuroscience*, 26(10):2738–2744, 2006.

[77] A. Camins, E. Verdaguer, J. Folch, and M. Pallàs. Involvement of calpain activation in neurodegenerative processes. *CNS Drug Reviews*, 12(2):135–148, 2006.

[78] C. Tarricone, R. Dhavan, J. Peng, L.B. Areces, L.H. Tsai, and A. Musacchio. Structure and regulation of the CDK5-p25nck5a complex. *Molecular Cell*, 8(3):657–669, 2001.

[79] M. Mapelli, L. Massimiliano, C. Crovace, M.A. Seeliger, L.H. Tsai, L. Meijer, and A. Musacchio. Mechanism of CDK5/p25 binding by CDK inhibitors. *Journal of Medicinal Chemistry*, 48(3):671–679, 2005.

[80] P.D. Jeffrey, A.A. Russo, K. Polyak, E. Gibbs, J. Hurwitz, J. Massagué, and N.P. Pavletich. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature*, 376(6538):313–320, 1995.

[81] M.A. Martí-Renom, A.C. Stuart, A. Fiser, R. Sánchez, F. Melo, and A. Sali. Comparative protein structure modeling of genes and genomes. *Annual review of biophysics and biomolecular structure*, 29(1):291–325, 2000.

[82] R.A. Laskowski, M.W. MacArthur, D.S. Moss, and J.M. Thornton. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2):283–291, 1993.

[83] G.N. Ramachandran, Ramakrishnan C., and Sasisekharan V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1):95–99, 1963.

[84] E. Lindahl, B. Hess, and D. van der Spoel. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, 7(8):306–317, 2001.

[85] W.L. Jorgensen, D.S. Maxwell, and J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.

[86] J.P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, 1977.

[87] W.G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, 1985.

[88] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981.

[89] W.G. Krebs and M. Gerstein. The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucleic Acids Research*, 28(8):1665, 2000.

[90] A. Laio, A. Rodriguez-Fortea, F.L. Gervasio, M. Ceccarelli, and M. Parrinello. Assessing the Accuracy of Metadynamics. *The Journal of Physical Chemistry B*, 109(14):6714–6721, 2005.

[91] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802, 2005.

[92] G. Henkelman, B.P. Uberuaga, and H. Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of Chemical Physics*, 113:9901, 2000.

[93] D.O. Morgan. Principles of CDK regulation. *Nature*, 374(6518):131–134, 1995.

[94] S.H.T. Saito. The regulation of cyclin-dependent kinase 5 activity through the metabolism of p35 or p39 Cdk5 activator. *Neurosignals*, 12:221–229, 2003.

[95] B. Nolen, S. Taylor, and G. Ghosh. Regulation of Protein Kinases: Controlling Activity through Activation Segment Conformation. *Molecular cell*, 15(5):661–675, 2004.

[96] E. Ozkirimli and C.B. Post. Src kinase activation: A switched electrostatic network. *Protein Science*, 15(5):1051–1062, 2006.

[97] E. Ozkirimli, S.S. Yadav, W.T. Miller, and C.B. Post. An electrostatic network and long-range regulation of Src kinases. *Protein Science*, 17(11):1871–1880, 2008.

[98] A. Cavalli, C. Dezi, G. Folkers, L. Scapozza, and M. Recanatini. Three-dimensional model of the cyclin-dependent kinase 1 (CDK1): Ab initio active site parameters for molecular dynamics studies of CDKS. *Proteins: Structure, Function, and Bioinformatics*, 45(4):478–485, 2001.

[99] S.Y. Wu, I. McNae, G. Kontopidis, S.J. McClue, C. McInnes, K.J. Stewart, S. Wang, D.I. Zheleva, H. Marriage, D.P. Lane, et al. Discovery of a Novel Family of CDK Inhibitors with the Program LIDAEUS: Structural Basis for Ligand-Induced Disordering of the Activation Loop. *Structure*, 11(4):399–410, 2003.

[100] S. Takahashi, T. Ohshima, A. Cho, T. Sreenath, M.J. Iadarola, H.C. Pant, Y. Kim, A.C. Nairn, R.O. Brady, P. Greengard, et al. Increased activity of cyclin-dependent kinase 5 leads to attenuation of cocaine-mediated dopamine signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 102(5):1737, 2005.

[101] A. Jämsä, A. Bäckström, E. Gustafsson, N. Dehvari, G. Hiller, R.F. Cowburn, and M. Vasänge. Glutamate treatment and p25 transfection increase Cdk5 mediated tau

phosphorylation in SH-SY5Y cells. *Biochemical and biophysical research communications*, 345(1):324–331, 2006.

[102] F. Sicheri, I. Moarefi, and J. Kuriyan. Crystal structure of the Src family tyrosine kinase Hck. *Nature*, 385(6617):602–609, 1997.

[103] W. Xu, S.C. Harrison, and M.J. Eck. Three-dimensional structure of the tyrosine kinase c-Src. *Nature*, 385(6617):595–602, 1997.

[104] T. Schindler, W. Bornmann, P. Pellicena, W.T. Miller, B. Clarkson, and J. Kuriyan. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science*, 289(5486):1938–1942, 2000.

[105] C. Tang, C.D. Schwieters, and G.M. Clore. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature*, 449(7165):1078–1082, 2007.

[106] A. Cuadrado and A. Nebreda. Mechanisms and functions of p38 MAPK signalling. *Biochemical Journal*, 429:403–417, 2010.

[107] S. Kumar, J. Boehm, and J.C. Lee. p38 MAP kinases: key signalling molecules as therapeutic targets for inflammatory diseases. *Nature Reviews Drug Discovery*, 2(9):717–726, 2003.

[108] C. Pargellis, L. Tong, L. Churchill, P.F. Cirillo, T. Gilmore, A.G. Graham, P.M. Grob, E.R. Hickey, N. Moss, S. Pav, et al. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nature Structural & Molecular Biology*, 9(4):268–272, 2002.

[109] R. Capdeville, E. Buchdunger, J. Zimmermann, and A. Matter. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nature Reviews Drug Discovery*, 1(7):493–502, 2002.

[110] J. Zhang, P.L. Yang, and N.S. Gray. Targeting cancer with small molecule kinase inhibitors. *Nature Reviews Cancer*, 9(1):28–39, 2009.

[111] A.C. Backes, B. Zech, B. Felber, B. Klebl, and G. Müller. Small-molecule inhibitors binding to protein kinases. Part I: exceptions from the traditional pharmacophore approach of type I inhibition. *Expert Opinion on Drug Discovery*, 3(12):1409–1425, 2008.

[112] A.C. Backes, B. Zech, B. Felber, B. Klebl, and G. Müller. Small-molecule inhibitors binding to protein kinase. Part II: the novel pharmacophore approach of type II and type III inhibition. *Expert Opinion on Drug Discovery*, 3(12):1427–1449, 2008.

[113] M. Vogtherr, K. Saxena, S. Hoelder, S. Grimme, M. Betz, U. Schieborr, B. Pescatore, M. Robin, L. Delarbre, T. Langer, et al. NMR Characterization of Kinase p38 Dynamics in Free and Ligand-Bound Forms. *Angewandte Chemie International Edition*, 45(6):993–997, 2006.

[114] Z. Wang, B.J. Canagarajah, J.C. Boehm, S. Kassisa, M.H. Cobb, P.R. Young, S. Abdel-Meguid, J.L. Adams, and E.J. Goldsmith. Structural basis of inhibitor selectivity in MAP kinases. *Structure*, 6(9):1117–1128, 1998.

[115] J.R. Simard, M. Getlik, C. Grütter, V. Pawar, S. Wulfert, M. Rabiller, and D. Rauh. Development of a fluorescent-tagged kinase assay system for the detection and characterization of allosteric kinase inhibitors. *Journal of the American Chemical Society*, 131(37):13286–13296, 2009.

[116] J. Regan, C.A. Pargellis, P.F. Cirillo, T. Gilmore, E.R. Hickey, G.W. Peet, A. Proto, A. Swinamer, and N. Moss. The kinetics of binding to p38 MAP kinase by analogues of BIRB 796. *Bioorganic & medicinal chemistry letters*, 13(18):3101–3104, 2003.

[117] T. Frembgen-Kesner and A.H. Elcock. Computational sampling of a cryptic drug binding site in a protein receptor: explicit solvent molecular dynamics and inhibitor docking to p38 MAP kinase. *Journal of molecular biology*, 359(1):202–214, 2006.

[118] F. Filomia, F. De Rienzo, and M.C. Menziani. Insights into mapk p38α dfg flip mechanism by accelerated molecular dynamics. *Bioorganic & Medicinal Chemistry*, 2010.

[119] Y. Shan, M.A. Seeliger, M.P. Eastwood, F. Frank, H. Xu, M.Ø. Jensen, R.O. Dror, J. Kuriyan, and D.E. Shaw. A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proceedings of the National Academy of Sciences*, 106(1):139, 2009.

[120] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.

[121] C.I. Bayly, P. Cieplak, W.D. Cornell, and P.A. Kollman. A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: the RESP model. *Journal of Physical Chemistry*, 97(10269-10280):279–280, 1993.

[122] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr, T. Vreven, K.N. Kudin, and J.C. Burant. Gaussian 03, Revision C. 02, Gaussian. *Inc., Wallingford, CT*, 5, 2004.

[123] W.D. Cornell, P. Cieplak, C.I. Bayly, and P.A. Kollmann. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *Journal of the American Chemical Society*, 115(21):9620–9631, 1993.

[124] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.

[125] M. Bonomi, A. Barducci, and M. Parrinello. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *Journal of Computational Chemistry*, 30(11):1615–1621, 2009.

[126] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia, et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications*, 180(10):1961–1972, 2009.

[127] E.M. Duffy, D.L. Severance, and W. L. Jorgensen. Urea: potential functions, log P, and free energy of hydration. *Israel Journal of Chemistry*, 3:323–330, 1993.

[128] L.J. Smith, H.J.C. Berendsen, and W.F. van Gunsteren. Computer Simulation of Urea- Water Mixtures: A Test of Force Field Parameters for Use in Biomolecular Simulation. *Journal of Physical Chemistry B*, 108(3):1065–1071, 2004.

[129] B. Hess, H. Bekker, H.J.C. Berendsen, and J.G.E. Fraaije. M.(1997) LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18:1463–1472.

[130] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *Journal of chemical physics*, 126:014101, 2007.

[131] A. Barducci, M. Bonomi, and M. Parrinello. Linking Well-Tempered Metadynamics Simulations with Experiments. *Biophysical journal*, 98(9):L44–L46, 2010.

[132] W.L. Jorgensen and J. Tirado-Rives. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.

[133] V. Munoz, P.A. Thompson, J. Hofrichter, and W.A. Eaton. Folding dynamics and mechanism of beta-hairpin formation. *Nature*, 390(6656):196–199, 1997.

[134] F.J. Blanco and L. Serrano. Folding of protein GB1 domain studied by the conformational characterization of fragments comprising its secondary structure elements. *European Journal of Biochemistry*, 230(2):634–649, 1995.

[135] K.K. Sinha and J.B. Udgaonkar. Dependence of the size of the initially collapsed form during the refolding of barstar on denaturant concentration: Evidence for a continuous transition. *Journal of molecular biology*, 353(3):704–718, 2005.

[136] U. Mayor, J. Günter Grossmann, N.W. Foster, S. Freund, and A.R. Fersht. The denatured state of engrailed homeodomain under denaturing and native conditions. *Journal of molecular biology*, 333(5):977–991, 2003.

[137] Y. Li, F. Picart, and D.P. Raleigh. Direct characterization of the folded, unfolded and urea-denatured states of the C-terminal domain of the ribosomal protein L9. *Journal of molecular biology*, 349(4):839–846, 2005.

[138] K.A. Merchant, R.B. Best, J.M. Louis, I.V. Gopich, and W.A. Eaton. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proceedings of the National Academy of Sciences*, 104(5):1528, 2007.

[139] P. Das and R. Zhou. Urea-induced drying of carbon nanotubes suggests existence of a dry globule-like transient state during chemical denaturation of Proteins. *Journal of Physical Chemistry B*, 114(16):5427–5430, 2010.