

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN
SCIENZE BIOCHIMICHE E BIOTECNOLOGICHE

Ciclo XXIX

Settore Concorsuale di afferenza: 03/D1

Settore Scientifico disciplinare: CHIM/08

TARGETING ALLOSTERIC POCKETS IN
PROTEIN KINASES USING MOLECULAR
MODELING AND SIMULATIONS

Presentata da: Giuseppina La Sala

Coordinatore Dottorato

Prof. Santi Mario Spampinato

Relatore

Dr. Marco De Vivo

Esame finale anno 2017

This thesis is dedicated to Pasquale

*“O frati,” dissi, “che per cento milia
perigli siete giunti a l’occidente,
a questa tanto picciola vigilia*

*d’i nostri sensi ch’è del rimanente
non vogliate negar l’esperienza,
di retro al sol, del mondo senza gente.*

*Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza”.*

Dante Alighieri
Inferno, canto XXVI, vv. 112-220

Acknowledgments

This thesis would have never seen the light without the help and the support of many people and I would like to use this opportunity to express my gratitude to them.

First and foremost, I am deeply indebted to my supervisor, Dr. Marco De Vivo. Thanks for giving me the opportunity to pursue my PhD under your guidance and for having taught me so much over these three years. Always with a positive outlook, you have constantly pushed me to go one step further or one level deeper with my research.

I would like to thank Prof. Matteo Dal Peraro, who kindly hosted me in his laboratory at EPFL for nine months. I was impressed by the fantastic environment that you created in your group. It was an inspiring place where I found high level researches, but also very good friends.

I am also immensely grateful to Prof. Oliver Hantschel, to Dr. Allan Lamontanara and to Daniel Duarte, who introduced me to the world of biophysics. Thanks to your suggestions and comments, I learnt how important it is to adopt different viewpoints to tackle a scientific question.

A special thank should be given to Dr. Walter Rocchia, Dr. Sergio Decherchi, Dr. Laura Riccardi and Dr. Roberto Gaspari, who analyzed and critically evaluated my work and my ideas. I benefited immensely from your deep insight and your constructive advice.

I would like to thank my colleagues Dr. Luca Mollica and Vito Genna. Coffee breaks and lunches would not be the same without you. You have always listened me in difficult times, giving me a laugh when I needed it.

I own a great debt of gratitude to my parents, Pietro and Antonella, and my sister Stefania for their love and unconditional support. You cannot imagine how rewarding it is to see you so proud of me. I hope you could consider my achievements, such as this doctorate degree, also a bit as your own.

My last thought goes to you, Davide. You have been my travelling companion during the last two years. Anyone like you has shared the with me the joys and the challenges of the PhD. You always supported and encouraged me. Definitely, this adventure would have been uphill without you by my side.

Contents

INTRODUCTION	11
<hr/>	
CHAPTER 1. ALLOSTERY IN PROTEIN KINASES	15
<hr/>	
1. ALLOSTERY IN PROTEINS	15
1.1. EXPERIMENTAL METHODS TO STUDY ALLOSTERY	18
1.2. COMPUTATIONAL METHODS TO STUDY ALLOSTERY	19
1.2.1. PROTEIN-SEQUENCE ANALYSIS METHODS	19
1.2.2. ENERGY LANDSCAPE	20
1.2.3. METHODS BASED ON MOLECULAR DYNAMICS SIMULATIONS	21
1.2.4. IDENTIFICATION OF PROTEIN ALLOSTERIC POCKETS	23
2. ALLOSTERY IN PROTEIN KINASES	25
2.1. PROTEIN KINASES: IMPORTANCE AND BIOLOGICAL FUNCTION	25
2.2. STRUCTURE OF PROTEIN KINASES AND ALLOSTERIC REGULATION	26
2.3. THE CASE OF ABELSON (ABL) KINASE	29
2.3.1. BIOLOGICAL FUNCTION AND CLINICAL SIGNIFICANCE	29
2.3.2. STRUCTURE AND ALLOSTERIC REGULATION	31
2.3.3. INHIBITION OF ABL	33
CHAPTER 2. THEORY	37
<hr/>	
1. DOCKING	37
2. MOLECULAR DYNAMICS SIMULATIONS	40
2.1. METADYNAMICS	42
2.2. ACCELERATED MD	45
CHAPTER 3. HRD MOTIF AS CENTRAL HUB OF THE SIGNALING NETWORK FOR ACTIVATION LOOP AUTOPHOSPHORYLATION IN ABL KINASE	49
<hr/>	
ABSTRACT	51
INTRODUCTION	52
RESULTS	55
DISCUSSION	65
METHODS	70
CHAPTER 4. AUTOMATED ANALYSIS OF BINDING POCKET DYNAMICS AND ALLOSTERIC NETWORKS IN PROTEINS	77
<hr/>	
ABSTRACT	79
INTRODUCTION	80
RESULTS	81
DISCUSSION	89
CONCLUSIONS	92
THEORY	93
METHODS	98

CHAPTER 5. STRUCTURAL CHARACTERIZATION OF BTK KINASE VIA INTEGRATIVE MODELING TECHNIQUES	101
INTRODUCTION	103
RESULTS AND DISCUSSION	104
METHODS	114
CHAPTER 6. DRUG DISCOVERY PROJECTS	119
1. RHOJ	119
1.1. INTRODUCTION	119
1.2. COMPUTATIONAL METHODS	120
1.3. PRELIMINARY RESULTS	121
2. ACID CERAMIDASE	123
2.1. INTRODUCTION	123
2.2. COMPUTATIONAL METHODS	124
2.3. RESULTS	124
APPENDIX A	127
APPENDIX B	147
BIBLIOGRAPHY	155
PUBLICATIONS	165
COMMUNICATIONS TO CONFERENCES	165

Introduction

Allostery is a regulation mechanism common to a number of proteins whereby a binding event of an effector molecule at one site affects another distinct functional (or active) site of the same protein. The effector can influence either the substrate affinity at the active site or the overall protein catalytic activity. For these reasons, the effector molecule can be either an allosteric activator or allosteric inhibitor, depending on whether it diminishes/blocks or increases protein function.¹

In the recent years, the development of allosteric modulators (both allosteric activators and allosteric inhibitors) has been on the rise because allosteric drugs offer a number of advantages that make them desired as drug candidates.^{2, 3}

Allosteric drugs, in fact, can exhibit a higher selectivity profile among proteins that belong to the same family than classical orthosteric drugs (i.e. drugs that bind the active site), because allosteric binding sites are often less evolutionarily conserved. Often, allosteric drugs are pursued to treat resistant forms of the diseases, because point mutations are often found in the orthosteric binding site, making the classical orthosteric drugs poorly effective, or even inactive. Allosteric drugs are also interesting from a physiological point of view. They provide a way to modulate the natural regulation (amplify or reduce naturally regulated response), rather than completely inhibit or activate a protein. In other words, the maximal effect of an allosteric binder does not necessarily lead to a complete protein inhibition or activation. This saturability allows safer doses. Overdoses of allosteric drugs, in fact, do not fully eliminate an essential signal. Also, allosteric ligands can have a spatiotemporal specificity. In practice, they can exhibit their function only in the presence of the endogenous ligand, restricting their effect to certain tissue at certain time.^{4, 5}

Notwithstanding their multiple advantages, the development of allosteric drugs is challenging. In fact, the number of allosteric drugs present in the market is

significantly lower respect to the orthosteric drugs.⁴ In this regard, one of the main issues encountered in the drug discovery process is the identification of the allosteric binding site on the protein surface. Very often, in fact, the allosteric binding pocket has a highly dynamical behavior, displaying a transient nature that makes it difficult to detect by means of conventional X-ray experiments.⁶ In addition, also an exhaustive knowledge of the molecular mechanisms that underlie the propagation of the allosteric signaling across the protein might help the process of the drug discovery.⁷ In this respect, allosteric drugs can be rationally designed in order to impede the functioning of the protein, working as a ‘stick wedged between the gears’ of the protein machinery. In this scenario, computational methods in conjunction with experimental techniques, as X-ray crystallography and NMR spectroscopy, might help both the detection of allosteric pockets and the comprehension of protein’s allosteric behavior.⁴

In this thesis, we applied computational techniques, based on molecular modeling and simulations, to face both the aforementioned aspects. First, we studied the molecular basis that allosterically regulates the function of Abelson (Abl) kinase, a relevant pharmaceutical target for the treatment of several malignancies, as chronic myelogenous leukemia (CML).⁸ This study proposes a novel mechanism according to which conserved structural motifs dynamically cooperate to regulate allosterically the function of the kinase. The information retrieved from this study can be employed for the rational design of new kinase allosteric inhibitors. In collaboration with Hantschel and Dal Peraro labs at École polytechnique fédérale de Lausanne (EPFL), where I spent a period of nine months during my PhD studies, we also investigated this mechanism in other kinases, as Btk, using an integrative modeling approach. In addition, we also developed a new algorithm for the detection of protein pockets in MD simulations. This algorithm has been conceived to identify and analyze all the pockets of a given protein without any user *a priori* information of their localization. It also enables the detection of pockets’ network, characterizing possible allosteric signaling pathways that connect the functional with the allosteric sites. Overall, this tool allows the study of the dynamic properties of pockets and might be employed in the early stages of the drug discovery process to design both orthosteric and allosteric binders. This study has been performed in collaboration with Concept Lab at IIT. Currently, this algorithm has been implemented in Pocketron, a module of BiKi software. To conclude, I gave my contribution in

several drug discovery projects. In this context, I employed computational methods to identify hit compounds able to inhibit validated targets for the treatment of cancer.

This thesis is divided in 7 chapters as follows:

- **Chapter 1:** General introduction on allostery in protein kinases, focusing on Abl, and on experimental and computational methods employed to elucidate allosteric mechanisms.
- **Chapter 2:** Description of the theoretical methods employed in this thesis (i.e., classical MD, enhanced sampling MD and molecular docking).
- **Chapter 3:** Study on the allosteric molecular mechanisms that control the A-loop plasticity in Abl kinase investigated via classical MD and free energy simulations.
- **Chapter 4:** Presentation of a new algorithm for the detection and analysis of protein pockets along extended MD simulation.
- **Chapter 5:** Ongoing work on structural characterization of activated Btk kinase structure investigated via integrative modelling.
- **Chapter 6:** Drug discovery projects this PhD course: application of virtual screening and molecular modeling techniques to identify inhibitors against pharmaceutical relevant protein target.

Chapter 1

Allostery in protein kinases

1. Allostery in proteins

Allostery is a regulation mechanism present in a number of biomolecular systems. In allostery, the binding of a small molecule (or a peptide or another protein) to a site far from the functional (or catalytic site) influences the overall protein activity. Allostery plays a fundamental role in a number of biological processes as signal transduction⁹, transcriptional regulation,¹⁰ molecular machines,¹¹ and metabolism,¹² with great relevance to disease.¹³ Given the paramount importance of allosteric processes in the regulation of cell life, it has been suggested that all proteins can be allosteric.¹⁴ In this context, the comprehension of protein allosteric mechanisms is essential in order to shed light on the functioning of biological processes and to explain the dysfunction of a disease to design drugs.¹⁵

Over the last decades, allostery has been largely studied and several models have been conceived in order to explain and rationalize the allosteric behavior in proteins. The first models proposed to describe the allosteric effects are the concerted MWC model by Monod, Wyman, and Changeux¹⁶ and the sequential KNF model by Koshland, Nemethy, and Filmer.¹⁷ Both models were originally developed to provide a description of allostery in hemoglobin. They consider two major conformational states for the different protein domains, the tensed (T) and the relaxed (R), where the relaxed binds the substrate more readily than the tensed. The MWC model assumes that enzyme subunits are in the same conformational state and are connected in such a way that a conformational change in one subunit is necessarily conferred to all other subunits. In absence of ligand, the equilibrium favors one of the conformational

state, which can be R or T. The binding of a ligand (allosteric effector) in a site different from the active site (allosteric site) shifts the chemical equilibrium in favor of the high-affinity state, leading to a cooperative ligand binding (Figure 1.1 A). On the contrary, the KNF model states that the enzyme subunits do not have necessarily the same conformation. The ligand induced fit binding leads to a conformational change in the domain, which is not propagated to the other subunits, as in the previous case. In fact, according to this model, the switch of one subunit makes the other subunits more likely to change, by reducing the energy needed for subsequent subunits to undergo the same conformational change (Figure 1.1 B).¹ However, the aforementioned models have in common the following features: i) only two protein states exist (i.e. the tensed or the relaxed); (ii) there is a conformational change in the substrate binding site; (iii) the allosteric signal from the allosteric site to the functional site is transmitted via a single pathway.¹⁸

Given the last advances in field of structural biology, proteins and other biomolecular systems acquire a new perception, in which they exist in a dynamical ensemble, rather than two discrete states. In this new scenario, a protein is seen from a statistic point of view, where different conformational states interconvert in one another, in an energy-driven manner. Therefore, the protein conformational ensemble has a certain distribution, which can be defined by the free energy landscape (Figure 1.1 C).^{19, 15} Notably, environmental changes as temperature, post-translational modification (PTM), point mutations, covalent or noncovalent ligand binding alter the protein population distribution. Such a state redistribution is the key of the allosteric effect. In this new view, allostery is a thermodynamic phenomenon, which is governed by enthalpy, enthalpy and entropy or entropy.^{15, 20} Moreover, the existence of multiple conformational and dynamic states implies the presence of multiple pathways in which the signal is transduced through the protein. The effect of an allosteric modulator can vary from major conformational rearrangements to subtle shift in conformational population. In this context, the allosteric signal can be transmitted through structural motifs, as rigid protein portions or flexible linkers and loops.⁴

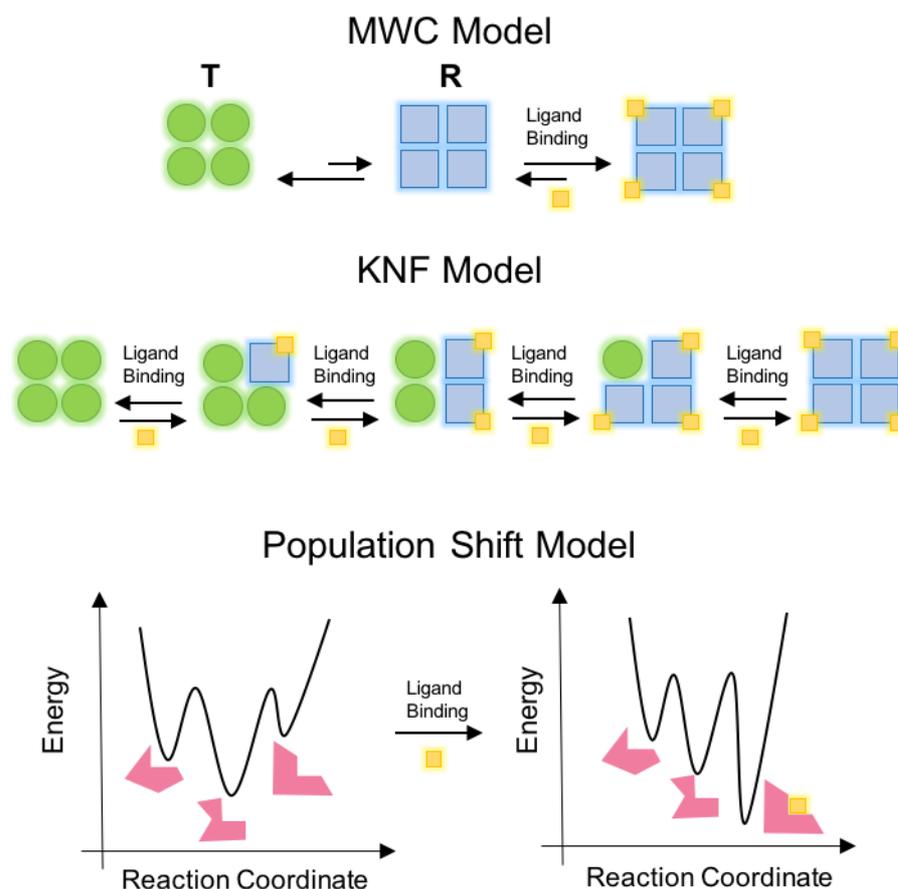


Figure 1.1. Models of allostery. (Top) The two-state concerted model of Monod–Wyman–Changeux (MWC) (Middle) The sequential model of Koshland–Nemethy–Filmer (KNF). (Bottom) The population shift model of allostery. Figure adapted from Lu S. et al.²¹

Albeit significant progress has been done to comprehend the principles that regulate allostery, it has to be investigated case-by-case. In fact, the molecular mechanisms that underlie the allosteric signal propagation can be different according to the molecular system that we are studying. Recently, it has been proposed that the allosteric signal can be propagated according two different models: the domino and the violin. In the domino model, the binding of an effector molecule to the allosteric site triggers local structural changes that are propagated linearly via a single pathway to the functional site.²² PDZ domain family²³ GPCRs, hemoglobin²⁴ are examples of proteins that possess this type mechanism. In the violin model, instead, the allosteric signal is propagated without a specific pathway, but it is transferred throughout the whole protein.²² Hereafter, we review the most employed experimental and computational methods to study allostery in protein, focusing on those that are also used to identify or predict allosteric binding sites.

1.1. Experimental methods to study allostery

During the last decades, protein allostery has been largely investigated by using experimental techniques. The X-ray crystallography is one of the most frequently used experimental approaches, providing a detailed characterization of static conformational states of a protein, for example before and after the allosteric signal. In fact, only after the structural characterization of both the tensed and the relaxed state of hemoglobin, Monod, Wyman, and Changeux were able to postulate the concerted MWC model.²⁵ However, since allostery is a dynamical process, the lack of dynamical information on the static crystal structures makes X-ray crystallography unsuitable to study alone the allosteric behavior of a protein, but requires the integration with other techniques. In this regard, nuclear magnetic resonance (NMR) spectroscopy delivers a more detailed information, capturing several transient conformations that can fill the gap among the protein states retrieved with X-ray crystallography. For example, through ¹⁹F NMR experiments, Ye and colleagues identified four new states of GPCR protein A_{2A} that helped the interpretation of the conformational landscape of the receptor. In this study the authors also found that these states, namely S1, S2, S3 and S3' are in equilibrium among them. This equilibrium can be perturbed in a ligand depend manner, according to a mechanism in line with the conformational selection theory. Importantly, the identification of new A_{2A} states enable to design potent and selective A_{2A} binders and the choice of the interacting state would then dictate pharmacology.²⁶

Also, the conformational changes of a protein can be captured with techniques as fluorescence resonance energy transfer (FRET) and hydrogen–deuterium exchange mass spectrometry (HDXMS). The former method measures the movements of two fluorophores properly attached to the protein, while the latter detects the replacement of hydrogens with the deuterium on the protein resulting from a conformation change. FRET was applied, for example, to study the mechanism of activation of APMA receptor. In particular, the authors of this study found the cleft closure in the ligand-binding domain of AMPA controls the activation of the receptor, in agreement with existing X-ray experiments.²⁷ Instead, HDXMS is an extensively used technique aimed at identifying conformational change in the protein upon ligand binding. Recently, Chandramohan et al successfully applied HDXMS

technique to map the dynamics of Hsp90, distinguishing the effects provoked by an orthosteric or an allosteric ligand.²⁸

1.2. Computational methods to study allostery

Computational methods play also a pivotal role in the study of protein allosteric mechanisms and in the identification of new targetable allosteric pockets. In this regard, molecular dynamics (MD) simulations (see Chapter 2) have emerged as a valuable complement to the experimental methods in the study of allostery.²⁹ MD simulations are particularly suitable for the study of allostery for three main reasons: i) they provide an all-atom description of protein behavior with very high resolution; ii) they allow to determine the effect of a perturbation, for example the introduction of a point mutation or the ligand binding, on protein's structure and motion; iii) they can be employed to find new allosteric pockets that are not visible in a ligand-free crystal structure.¹⁴ Hereafter, we report the main computational approaches that are used to probe the mechanistic aspects of allostery and to detect allosteric pockets in proteins.

1.2.1. Protein-sequence analysis methods

The protein-sequence analysis is a powerful bioinformatics-based method that enables the identification of allosteric pathways as well as protein pockets. This method is based on the premise that a residue conserved during the evolution is likely to have a functional role.²³ To perform this analysis, one needs a query sequence (i.e. the protein that one want to study) and a set of sequences of homologues proteins. After the alignment of the sequences to the query, a specific algorithm is applied in order to return evolutionary important residues. Depending on the type of algorithm used, this analysis returns two types of outcome. The first one is a list of single sequence positions without predicting a specific connection between them. In other words, the analysis predicts evolutionary important residues across the protein, but does not indicate whether far residues are linked to form the same allosteric pathway or not. Conversely, the second outcome returns pairs of evolutionary important residues that are linked between them.³⁰ To do so, the algorithm detects the residue pairs that mutate in concert more frequently than one

would expect given random genetic events. In this sense, coevolved residues are assumed to be part of the same allosteric pathway. Despite these analyses are appealing because of their simplicity, they suffer of some limitations. Firstly, the specific biological role of the predicted evolutionary important residues is not indicated. This means that one cannot determine without an additional knowledge if a given residue might have catalytic role, might be part of a pocket, might have a structural role or might have an allosteric role. Another drawback of this method is that the results are strongly dependent on the goodness of alignment and on the type and number of structure aligned.³¹ Protein-sequence analysis was employed in different biomolecular systems, identifying new allosteric pockets as well as functional residues important for the allosteric signal transmission. For example, co-evolutionary analysis was successfully used to study the allostery in PDZ domain family.²³

1.2.2. Energy landscape

In statistical mechanics, the free energy landscape of a biomolecular system is a function of the probability of observing a particular event. The deeper is the basin of a particular configuration, the more likely the protein visits that particular state. Also, the transition among different states is more probable if the energy barriers that divide the states are low. In this context, a positive allosteric modulator modifies the energy landscape of the unbound protein, increasing the depth of the binding site well or decreasing the depth of the competitive wells.¹⁵ Here, we report several methods that take advantage of the principles based on energy landscape to understand allostery in proteins and to identify ligand binding sites.

The first method is based on protein frustration. This method starts from the assumption that proteins are endowed of a precise set of residues that can be key in the transmission of the allosteric signal or can be functional for ligand binding. These residues are referred to as “frustrated” and can be identified through heuristic algorithms.³² Ferreiro et al. for example, systematically perturbed the protein by mutating the residues and comparing the total energy of the mutated protein with the wild type. If the mutated protein is more energetically favored than the wild type, then the residues that have been mutated are considered frustrated.³³ Weinkam and

colleagues successfully employed protein frustration in CLS transcription factor, obtaining useful insights into allosteric mechanism of the protein.³⁴

The Normal-Mode Analysis (NMA) is another widely-used method to study protein allostery. NMA examines the vibrational motion of a harmonic oscillating system in the immediate vicinity of its equilibrium. In proteins, NMA is based on the assumption that vibrational normal modes with the lowest frequencies describe the largest movements in a protein and are the ones that are functionally relevant.³⁵ The binding of an allosteric effector perturbs the system, yielding new fundamental modes that might provide insights into allosteric mechanisms. NMA was successfully employed in several studies. By means of NMA, in fact, Gaillard et al. were able to have useful insights on the conformational transition between the high affinity ligand and low affinity ligand in LFA-1 protein.³⁶ Also, NMA can be used to identify residues in the protein surface that can form a ligand orthosteric or allosteric binding site. In this regard, AlloPred is a tool developed recently by Greener et al that coupled NMA and machine learning techniques to predict protein binding sites.³⁷

An alternate option to NMA is the Elastic Network Model (ENM). In ENMs models a uniform harmonic potential is used to model the interactions between residue pairs. In this scenario, protein is seen as a 3D network, where the edges are virtual springs with a given potential energy. ENM was used to study global dynamics and allostery in a number of biomolecular structures, such as enzymes,^{38 39} ribosome⁴⁰ and viral capsids.^{41 42}

1.2.3. Methods based on molecular dynamics simulations

Allostery is regulation mechanism that involves protein dynamics. In particular, an allosteric effector perturbs the protein function either through large-scale structural changes (i.e. domains conformational changes, lid domain motions, loop rearrangements) or through subtle residues motions.⁷ In this respect, molecular dynamics (MD) simulations represent an effective tool to extensively study allosteric behavior of proteins, because of their ability to capture protein fluctuations and motions. So far, a wealth of computational studies adopted MD simulations to investigate allostery in proteins. These studies achieved useful insights on molecular mechanisms that underlie the allosteric behavior of a protein and were also able to identify binding sites of allosteric effectors. To study allostery several methods,

based on MD simulations, exist. Among them, the most used are methods based on the network representation (or graph-based representation) of the protein and community analysis.

The graph-based analysis is one of the most employed computational approach to probe the mechanistic aspects of allostery, employing information from MD simulation. According to this analysis, the 3D structure of a protein is represented as a graph where the nodes correspond to individual residues.¹⁴ Each pair of nodes is connected by an edge whose length is inversely proportional to the degree of interdependence between their motions. The “interdependence” between nodes can be computed with different methods that are based on correlated motions (e.g., mutual information,⁴³ fluctuation in atomic positions⁴⁴ or on the number of specific noncovalent interactions (Figure 1.2). At this point, the resulting network representation of the protein dynamics can be analyzed with several techniques, in order to retrieve important pathways of communication between distant residues that contribute to allostery.⁴ Recently, Rehn et al. focused on the correlated motions, based on atomic fluctuation, to study the allosteric activation and deactivation of Hsp90. Supported by experimental evidences, the authors identified key residues, called ‘hot spots’, that are responsible of activating and inhibiting the allosteric signal in Hsp90.⁴⁵

Communities of residues (i.e. group of nodes in the network) that have a high interdependence among them can be considered an isolated module of the protein. In this regard, protein allostery can be seen as a dynamical connection among different communities that cooperate for the transmission of the allosteric signal (Figure 1.2). Using this approach, Rivalta et al. obtained useful insights on the allosteric mechanism of PRFAR binding to IGPS heterodimer, which were also supported by NMR experiments.⁴⁶

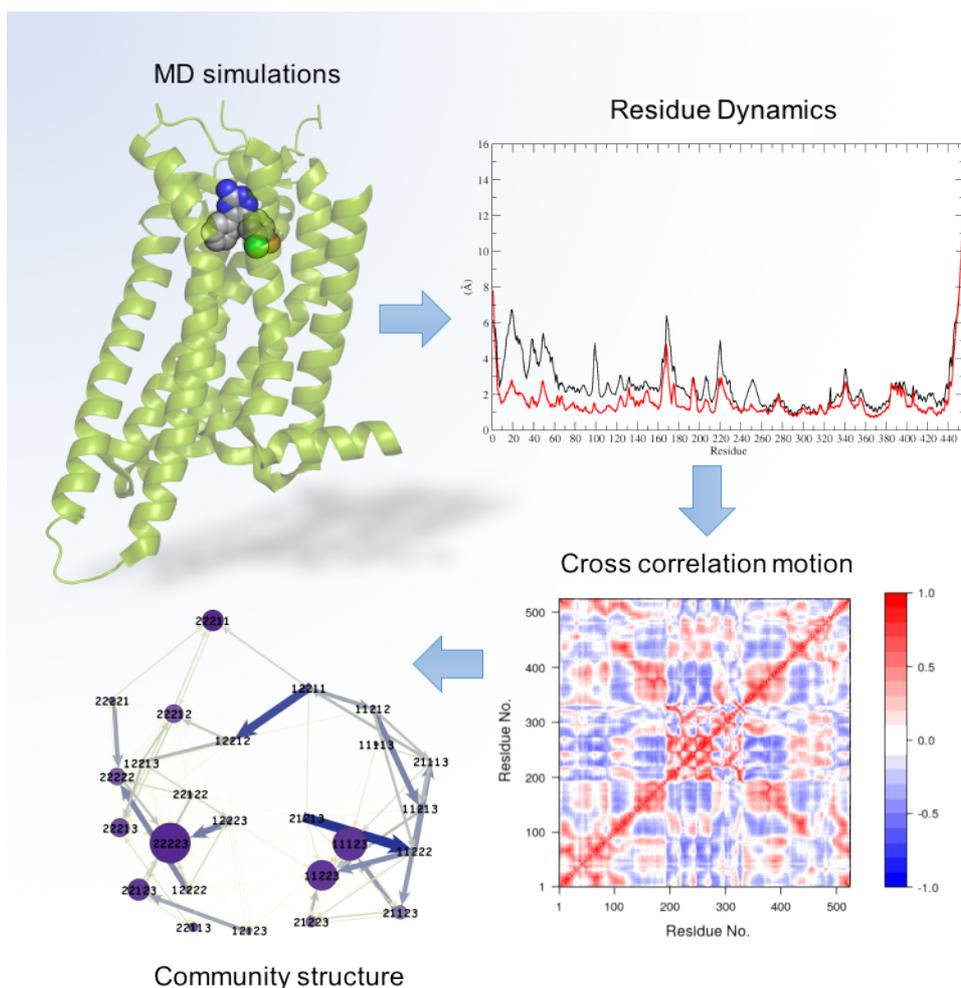


Figure 1.2. Illustration of the main approach for building community structure, using GPCR as an example. (Upper left) GPCR cartoon representation. (Upper right) Degree of per residue backbone dynamics derived from molecular dynamics simulations. (Bottom right) Example of matrix of residue-residue cross-correlated motion derived from molecular dynamics simulations. (Bottom left) Communities derived from the cross-correlation matrix represented as nodes. Inter-community edge widths (proportional to the cumulative betweenness of intercommunity edges), represent the strength of the potential allosteric communication between communities. Figure adapted from Schueler-Furman, O. and Wodak, S.J.¹⁴

1.2.4. Identification of protein allosteric pockets

Binding pockets are cavities on the protein surface where a drug interacts to perform its function. Nowadays, a number of algorithms and methods are present, which aimed at detecting protein pockets and determine their physicochemical characteristics.⁴⁷ The major part of these tools relies on static structures that belong

to X-ray, NMR experiments or selected frames from MD trajectories. These pocket detection methods can be classified as geometry, knowledge and energy based.

- *Geometry-based methods.* These methods identify pockets by considering only the 3D spatial position of atoms within a protein. These methods return important information as volume and pocket's shape, neglecting however other insights as the pocket physicochemical properties. In turn, geometric-based methods can be also classified in other three categories: sphere-based, the α shape-based and the grid-based methods. In sphere-based methods, a number of spheres are placed around the protein. All the spheres are subsequently removed except those that occupy a surface cavity, which form therefore a binding pocket. The α shape can be considered a generalization of a convex hull, although it is more precise respect to a convex hull. A number of methods employ this theory to define a pocket. Lastly, in grid-based methods the protein immersed in a grid and the grid points that lie a pocket surface cavity are clustered to form a pocket.⁴⁸
- *Knowledge-based methods.* These methods combine information that belongs to structural and genomic databases, taking into account of sequence and structural conservation to predict the position of a protein binding pocket.⁴⁸
- *Energy-based methods.* These methods compute the energy between the protein surface residues and a probe, which in most cases is an organic molecule as methane or a water molecule, and identify binding site according to the resulting binding affinity.⁴⁸

Despite the growing number of accurate and robust algorithms for the detection of pockets, the identification of allosteric pockets is still challenging. In fact, quite often they are not detectable by looking only static structures. Frequently, allosteric pockets are transient and can be formed only after a conformational change that classical X-ray or NMR experiments are not able to capture. In this regard, coupling pocket detection algorithms with MD simulations might represent a viable strategy to overcome this issue. In the last years, many efforts have been conducting to develop new methods for the study of pocket behavior in a dynamic context. In 2011 Schmidtke et al. developed the first tool able to pocket dynamics from an MD-generated ensemble.⁴⁹ Since then, other groups conceived different algorithms aimed at efficiently retrieving important information as pocket appearance, pocket volume

trend, pocket persistency along time. In this regard, in collaboration with the Concept lab in IIT, we also developed a novel algorithm for the identification and the analysis of protein pockets along extended MD simulations. Unlike the already existing algorithm, our tool also allows to determine whether near and far pockets communicate among each other. This return a network-based representation of protein pockets from which one can retrieve pathways that mediate the allosteric communication between functional and allosteric site (see Chapter 4).

2. Allostery in protein kinases

2.1. Protein kinases: importance and biological function

Protein kinases (PKs) constitute nearly the 2% of human genome⁵⁰ and are one of the most important protein families, because they regulate a number of biological process, ensuring the correct cell functionality. PKs catalyze the transfer of γ -phosphate from ATP to specific substrates as serines, threonines, tyrosines or histidines belonging to peptides or proteins. PKs phosphorylate about a third of all proteins in human cells.⁵¹ The phosphorylation causes significant changes on the structures of substrate proteins and affects also their function, activating or inhibiting their activity.^{52, 53} Therefore, PKs are considered one of the main regulators of cell biology, tightly controlling a number of signaling pathways important for cell growth and survival. Given their paramount importance in many biological processes, the activity of PKs is tightly regulated through several mechanisms. The lack or the impairment of such regulation mechanisms cause the dysfunction of PKs, which in turn leads to the incoming of multiple diseases as cancer, vascular disorders, diabetes, as well as neurological and inflammatory diseases. This makes protein kinases attractive pharmaceutical targets. In fact, the inhibition or the modulation of PKs activity became a validated strategy for the treatment of several diseases. So far, nearly 30 small-molecule protein kinase inhibitors are approved by the US FDA, and thousands of clinical studies targeting kinases are ongoing.^{54, 22}

2.2. Structure of protein kinases and allosteric regulation

Eukaryotic protein kinases (EPKs) are endowed of a kinase domain (KD) of approximately 250 amino acids that has a catalytic role. Depending on the kinase family, KD can be coupled with other domains both at the N- and an C-terminal region. These domains can have a regulatory function, modulating the kinase activity or can mediate the interactions with other proteins, DNA or with the cellular membrane. For example, Src, Csk, Brk and Abl kinase families are endowed of SH3 and SH2 domains, which play a key role in the allosteric modulation of kinase function and in the protein- protein interaction. Instead, Tec and Fes family kinases have the PH and the F-BAR domain, respectively, that allow the anchoring with the membrane interacting with the phospholipid. However, the KD is the most conserved region across all the kinases families. It has a bilobed shape formed by a small N-lobe a larger C-lobe. The N-lobe is formed by five antiparallel β -sheets and a single α -helix named α C, while the C-lobe is chiefly formed by α -helices. The N- and the C-lobe are connected through a short sequence of residues called 'hinge region'. This portion forms the ATP binding site, which is the orthosteric pocket that interacts the ATP and kinase ATP-competitive inhibitors. The ATP binding site is also formed by the G-loop, a glycine-rich sequence between β 1 and β 2 strands of the N-lobe that interacts with the phosphates of ATP, coordinating the ATP-Mg²⁺ co-substrate complex. The C-lobe also contributes to the ATP binding site through the activation loop (A-loop). This is a structural feature common in kinases, which is formed by a sequence of ~30 amino acids delimited by the DFG and the APE motifs in the N-terminal and in the C-terminal part, respectively. The DFG motif has a catalytic role because it directly coordinates the Mg²⁺ ions, while the C-terminal part of the A-loop interacts and positions the substrate (see Figure 1.3).⁵⁵

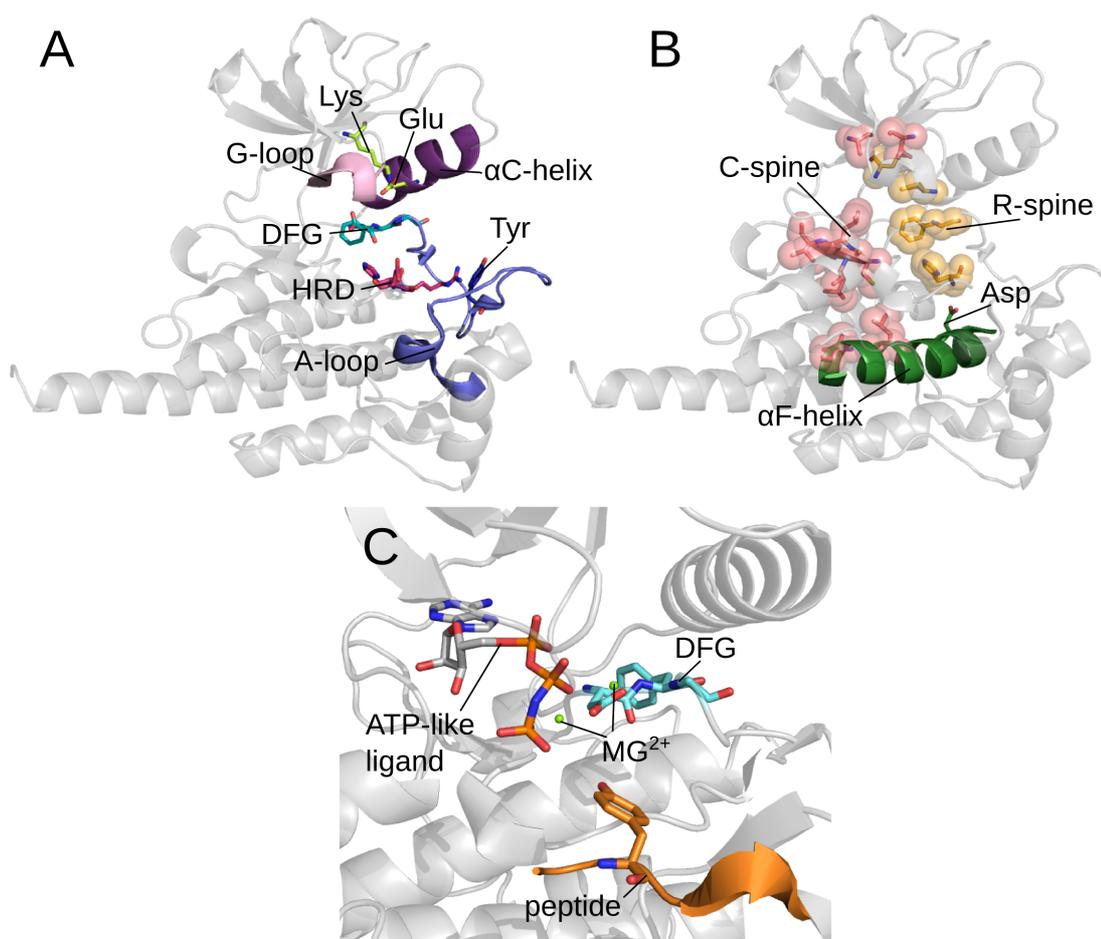


Figure 1.3. A) Kinase domain of Abl kinase in the active conformation. PDB code 2F4J. Important catalytic elements discussed in the text are reported. B) Representation of the C-spine (red) and R-spine (orange) in Abl kinase. Both spines are anchored to the α F-helix (green). PDB code 2F4J. C) ATP binding site of Insulin receptor kinase in complex with ATP-like molecule (grey sticks) and a peptide (orange). PDB code 1IR3.

So far, X-ray and NMR experiments captured protein kinases in several active, inactive and intermediate conformations. Importantly, it turned out that when active, protein kinases share common structural organization, albeit they belong to different families.²² Conversely, when inactive, protein kinases adopt a particular structural assembly that is family-dependent.⁵⁶ In particular, in the active state catalytic important structural elements of KD, as the α C and the A-loop, are proper oriented and constrained to foster the phosphotransfer reaction. Here, the DFG motif is found in the so called *in* conformation, where the phenylalanine lies in the ATP binding site and the aspartate is in an optimal orientation to coordinate the Mg^{2+} ions. The α C helix is rotated toward the ATP binding site enabling the formation of the conserved

salt bridge between the lysine of β 3-sheet and the glutamate of the α C helix. The A-loop assumes an open conformation, enabling both the ATP and the substrate entrance. In the major part of protein kinases, the phosphorylation of A-loop at tyrosine, serine or threonine residues, further stabilizes the open conformation of this segment. This is made possible through the formation of a salt bridge between the phosphate and the arginine located in the conserved HRD motif. Recently, it has been found that kinases share two hydrophobic motifs, whose proper alignment is related to the active kinase state. The two motifs are the regulatory spine (R-spine) and the catalytic spine (C-spine) which are both spines are anchored to the hydrophobic α F-helix.^{57, 58}

From a functional point of view, protein kinases are molecular regulators in many biological processes, activating or inactivating specific cellular pathways. To efficiently perform their function, kinases have evolved fine allosteric regulation mechanisms that enable them to promptly switch on or off in response to external agents. This ‘switch-like’ feature has been made possible thanks to their structural plasticity and their dynamical architecture. In particular, kinases can be turned on or turned off through four main mechanisms:

- 1) *Interaction with regulatory subunits.* Together with KD, a number of kinases are endowed of regulatory subunits that allosterically modulate the catalytic function in KD. A paradigmatic example of this class of mechanism is represented by PKA kinase. PKA is a tetrameric enzyme, which is formed by two KDs and two regulatory subunits named CNB domains, which in turn are divided in two subunits, CNB-A and CNB-B. In absence of the second messenger cAMP, the CNB domains bind the KDs, suppressing their activity. The binding of four molecules of cAMP to the two CNB domains enable a conformational change in CBNs that allow the exposure of KDs;⁵⁹
- 2) *Autoinhibition.* This is a mechanism of autoregulation found for example in Ca^{2+} -regulated kinases, twitchin⁶⁰ and titin.⁶¹ The KD of these protein kinases are endowed of a C-terminal regulatory motif that acts as pseudosubstrate competing directly with the substrate entrance to the active site. When the C-terminal tail enters in the ATP binding site, it causes a rotation of the α C helix that leads to the disruption of the catalytic important Lys-Glu salt bridge (Lys82-Glu98 in twitchin and Lys55-Glu68 in titin);

- 3) *Dimerization*. This is mechanism by which the kinase is activated via the binding of another kinase. A paradigmatic example of this mechanism is represented by EGF kinase. In normal conditions, the kinase is inactivated displaying the α C helix rotated outward the ATP binding site and the A-loop folded. The binding of another EGF kinase (called activator kinase) forms the so-called head-to-tail dimer. This leads to the activation of the receiving kinase through the stabilization of the open A-loop conformation and the rotation of the α C helix.⁶²
- 4) *Phosphorylation by another kinase (trans-phosphorylation) or by itself (cis-phosphorylation)*. Phosphorylation is one of the most common mechanism employed by protein kinases to enhance or inhibit their function. Phosphorylation on particular residues, in fact, drastically change the overall charge of the protein leading to conformational changes stabilizing the active or the inactive form of the protein. An example of this is provided by Src kinase. The phosphorylation of the Y527 at the C-term stabilizes the closed autoinhibited conformation, where the phosphate docks a pocket in the SH2 domain. Conversely, the phosphorylation of the A-loop at Y416, stabilizes the active conformation of the kinase.⁵⁵

2.3. The case of Abelson (Abl) kinase

2.3.1. Biological function and clinical significance

In this thesis, we mainly focused on Abelson (Abl) tyrosine kinases, a cytoplasmic and nuclear protein encoded by *ABL1* gene located in chromosome 9. Abl plays an important role in several signaling pathways that control cell differentiation, cell division, cell adhesion and stress response. Mutations in the *ABL1* gene often lead to the deregulation of Abl activity, which in turn results in the development of diverse pathologies, including several solid tumors, inflammatory disorders and neurodegenerative diseases.⁶³

The most common Abl-related disorder is the chronic myelogenous leukemia (CML), a cancer of white blood cells characterized by the unregulated growth of myeloid cells in the bone marrow and by the accumulation of these cells in the blood. The incoming of CML is driven by the chromosomal translocation of *ABL1* gene in

chromosome 22 within the *BCR* (breakpoint cluster region) gene, generating the so-called Philadelphia chromosome (Ph). The resulting fusion gene *BCR-ABL* encodes the unregulated Abl protein, the Bcr-Abl form, enabling the uncontrolled proliferation of cancerous cells in blood⁶⁴ (Figure 1.4).

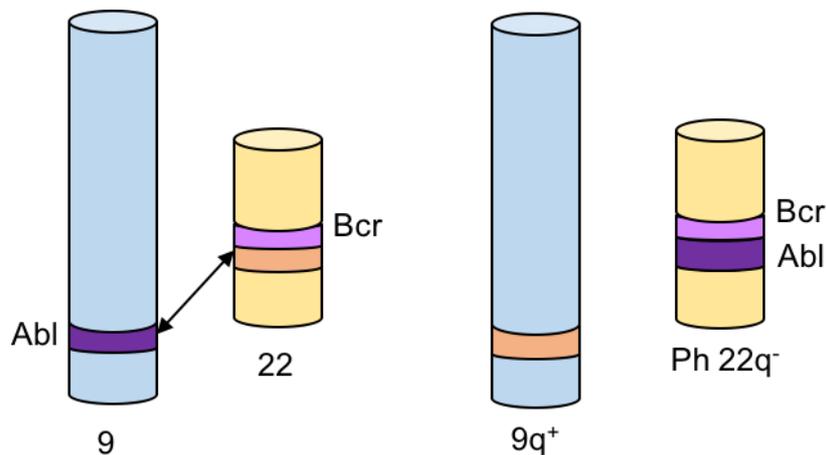


Figure 1.4. The *ABL* and *BCR* genes reside on chromosomes 9 and 22, respectively. As a result of the (9;22) translocation, a *BCR-ABL* fusion gene is formed on the derivative chromosome 22 (Philadelphia chromosome).

The inhibition of Abl activity, through ATP-competitive small molecules, is a viable strategy for the treatment of CML.⁶⁵ In 2001 the Food and Drug Administration (FDA) approved the first Abl inhibitor, Imatinib (or known also as Gleevec), a blockbuster drug that induce durable remissions and progression free survival in the majority of chronic-phase Ph-positive CML patients⁶³ (see paragraph 2.3.3.). However, over the last years, the first line therapeutics became ineffective against CML because of the incoming of mutated forms of Abl, which make current drugs poorly or completely inactive. As a consequence, nowadays many efforts are doing in order to design potent and selective Abl inhibitors that are also able to overcome the Imatinib-resistant forms of Abl.⁶⁶ In this respect, one of the main strategy consists of targeting Abl allosteric pockets, which are binding site far from the functional ATP binding site. Structural and biochemical studies, in fact, have demonstrated that allosteric binders are key in the modulation of Abl activity stabilizing the active or the inactive form of the enzyme. In this context, the fully

comprehension of molecular mechanisms at atomistic level that underlie allosteric regulation in Abl is required to help the rational design of Abl allosteric inhibitors.

2.3.2. Structure and allosteric regulation

Abl kinase is a 123 kDa protein that consists of ~1150 residues. Abl has a modular structure formed by the N-terminal cap, the SH3 and the SH2 domains and the kinase domain (KD) and, finally, by a series of binding elements for SH3 domain, a DNA binding functionality and an actin binding domain. In human, Abl is expressed in two isoforms: the Abl 1a and the Abl 1b, which differ only for the N-terminal part. In particular, the 1b isoform is myristoylated, while the 1a is not (Figure 1.5).⁶⁷

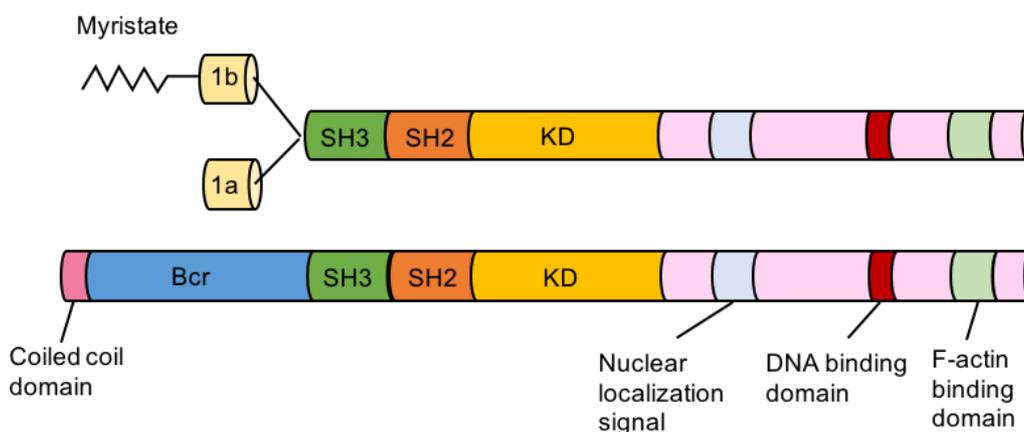


Figure 1.5. Domain structures of Abl kinase on the top (both isoforms 1b and 1a are represented) and of the oncogenic Bcr-Abl kinase on the bottom.

The activity of Abl kinase is regulated by a trans-phosphorylation mechanism. According to this mechanism, another kinase (as another Abl kinase itself) phosphorylates several residues spanned across the kinase structure, provoking a conformational rearrangement that ends up with the stabilization of the catalytic active form of Abl. In fact, when not phosphorylated, Abl assumes an inactive (autoinhibited or closed) conformation where the SH3 and the SH2 domains pack against the lateral part of KD (Figure 1.6).^{68, 69} In contrast, when phosphorylated, Abl is in an activated state where the SH2 binds the N-lobe of KD, assuming an elongated conformation (Figure 1.6). In this state, important catalytic elements of

KD as the α C and the A-loop, are properly oriented and constrained to foster the phosphotransfer reaction (see paragraph 2.2.).^{69, 70}

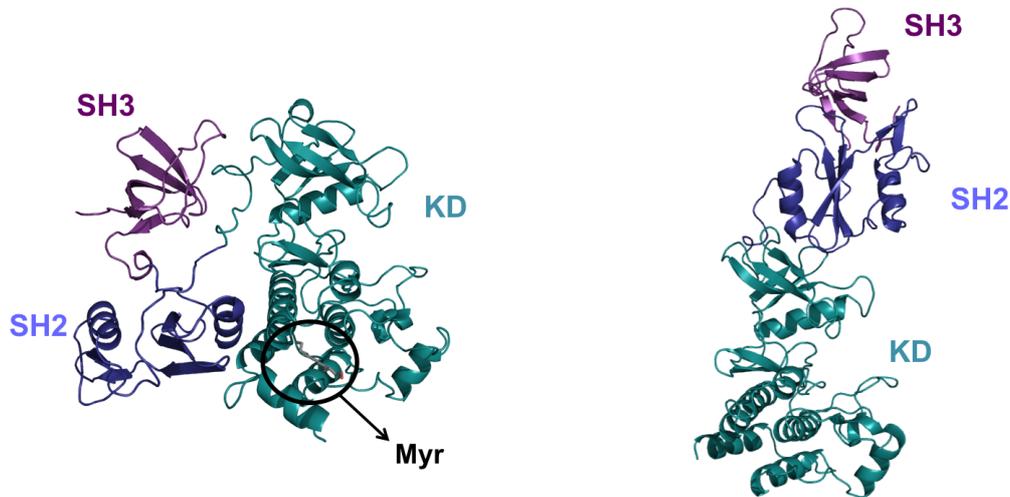


Figure 1.6. Representation of the two main conformations of Abl kinase. On the left the autoinhibited (closed) conformation is depicted (PDB code: 1OPK). The SH2 and the SH3 domains bind the lateral part of KD, while the myristate (Myr) interacts with the C-lobe of KD. Because of the low resolution of the crystal the N-cap is not shown. On the right, instead, the activated (elongated) conformation is shown (PDB code: 1OPL chain B). The SH2 domain binds the N-lobe of KD. The original X-ray structure comprehends only the SH2-KD complex. For sake of clarity the SH3 domain was modeled.

The N-cap is a region of ~80 amino acids which is important for kinase autoinhibition. The myristic group, present only in the 1b isoform,⁷¹ binds a deep hydrophobic pocket in the C-lobe of KD (Figure 1.6), which is a critical step for the maintenance of the autoinhibited form of Abl.⁷² In this way, the myristate acts as allosteric modulator of Abl function, tightly regulating its activity. In fact, the deletion of this region as well as the presence of point mutations result in a highly active kinase.⁷² This is the case of the Bcr-Abl oncogenic protein, where the N-cap region is replaced with the Bcr protein (Fig. 3). The SH3 domain is formed by two antiparallel β -sheets packed forming a barrel-shaped structure. The SH3 domain binds the proline-rich peptide that adopt a polyproline type II (PPII) conformation.⁷³ When in the autoinhibited Abl form, the SH3 interacts with the SH2-KD linker further stabilizing the closed conformation of Abl. Deletion of SH3 domain leads to an up-regulation of Abl activity, further confirming the importance of SH3 in the negative regulation of Abl.⁷⁴ The SH2 domain is formed by a central antiparallel β -

sheet flanked by α -helices on each side. The SH2 binds specific phosphotyrosine sequences, contributing to recognition and to phosphorylation of Abl substrates. This binding site is formed by a conserved FLVRES motif which interacts with the phosphotyrosine residue and by a nearby pocket that interact with the third amino acid of the C-terminal conferring substrate sequence specificity.⁷⁵ When in the autoinhibited form, the SH2 docks the C-lobe of KD via a network of H-bonds and a unique pi-stacking involving Tyr 158 of SH2 and Tyr 361 of KD. Notably, this interaction is favored by a conformational rearrangement of the α I helix of KD which fold of $\sim 45^\circ$ after the binding of myristate (or another allosteric inhibitor, see later) to the C-lobe pocket in KD. Mutations in the SH2/C-lobe interface increases the Abl activity, evidencing the importance of SH2 in the down-regulation of Abl.⁷² When in the activated form, the SH2 domain binds the N-lobe of KD. Interestingly, this interaction has been found to be important for stabilizing the active conformation of Abl (see Chapter 4 for further details), augmenting also the accessibility of the A-loop to the substrate. In fact, biochemical studies have demonstrated that the SH2-KD construct presents a higher catalytic activity than KD alone. Additionally, the SH2-KD complex favor the opening of the A-loop and the subsequent phosphorylation at Y412, which is a required step for a full kinase activation. Mutational studies also confirmed the importance of this domain-domain interaction in the allosteric modulation of Abl activity. In fact, point mutations in SH2 domain that impair the binding (e.g. I164E) are linked with a decreased activity of Abl,⁷⁶ whereas mutations that strengthen this interaction (e.g. T231R) lead to an increased Abl activity.

2.3.3. Inhibition of Abl

Drugs that inhibit the oncogenic form of Abl kinase (Bcr-Abl) currently in the market bind the orthosteric pocket at the ATP binding site. On the basis of their molecular mechanism of action, the Bcr-Abl orthosteric inhibitors can be distinguished in two classes, the type 1 and the type 2. The type 1 inhibitors interact with the active conformation of the kinase, where the catalytically important residues are in the optimal position to foster the catalysis. This means that the DFG in the *in* conformation and the A-loop is open. On the contrary the type 2 inhibitors bind the inactive conformation of the Bcr-Abl. In this context, the DFG motif in the *out*,

while the A-loop can be found either in the open or in the closed conformation. Imatinib is the first Bcr-Abl inhibitor that entered in the market in 2001 and, from then on it is still the first line drug to the treatment of CML.⁷⁷ Imatinib is a type 2 inhibitor, showing a high selectivity profile and a IC_{50} of 10.8 nM (Figure 1.7). However, the incoming of mutations spanned across the Abl protein, makes Imatinib poor or completely inactive against certain mutated form of the enzyme. The most common Imatinib-resistant mutations are: G250E, Q252H, Y253H/F, E255K/V, T315I, H395P/R, M244V, F317L, M351T, E355G, and F359V. The second generation of Bcr-Abl inhibitors is active against all the mutated form of the enzyme with the exception of the T315I form. Dasatinib and Nilotibib are examples of this class of inhibitors (Figure 1.7). Dasatinib is a type 1 inhibitor, having an IC_{50} <0.2 nM, while Nilotinib is a type 2 inhibitor with an IC_{50} of 38 nM. In 2012, the FDA approved the first drug belonging to the third generation of Bcr-Abl inhibitors, which are active against the T315I mutation. This drug, Ponatinib (Figure 1.7), has been developed by ARIAD pharmaceuticals and has a IC_{50} 1 nM against the T315I Bcr-Abl.⁷⁸

Together with the ATP orthosteric pocket, structural and biochemical investigations have found that also the pocket at the myristate binding site can be target to inhibit the Abl function. Compounds as GNF-2 (Figure 1.7) and GNF-5 are, in fact, able to inhibit Bcr-Abl-dependent cell growth and different Imatinib resistant mutations with exception for T315I. However, the combination of GNF-2 with Nilotinib has shown to have a synergistic effect, being able to inhibit also the T315I form. Therefore, the combination of orthosteric and allosteric inhibitors represents an innovative way to overcome resistance to either agent alone.

Finally, also the impairment of the SH2/N-lobe interface has been found a viable alternative to inhibit allosterically the function of Abl kinase. Recent studies have demonstrated that monobodies with highly affinity of the SH2 prevent the formation of the SH2-KD complex. This induces the apoptosis in CML cell lines and primary human CML cells.⁶⁶

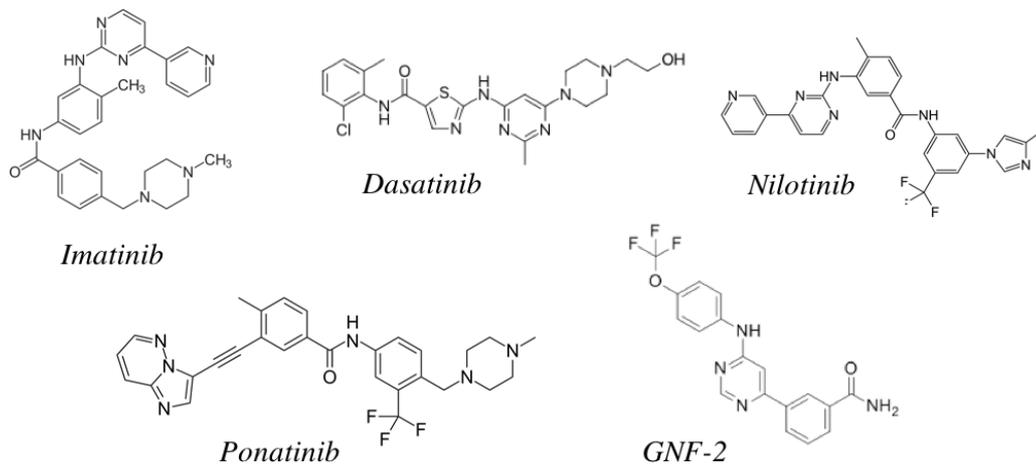


Figure 1.7. 2D structures of five *Abl* inhibitors

Chapter 2

Theory

1. Docking

Molecular docking is a computational technique to predict the binding orientation of one ligand into a biomolecular target (i.e. receptor) to form a stable complex (Figure 2.1 A). Such information, therefore, might be used to compute the strength of association as well as the stability of a given complex. This can be done employing scoring function of molecular docking.⁷⁹

Docking can be achieved via two steps: i) sampling the conformations of the ligand in the active site of the receptor (i.e. posing step); ii) computing the affinities between each ligand's conformation and the receptor (i.e. scoring step). Ideally, sampling algorithms should reproduce the experimental binding mode, while scoring function should discern between inactive and active molecules, ranking them highest.⁸⁰

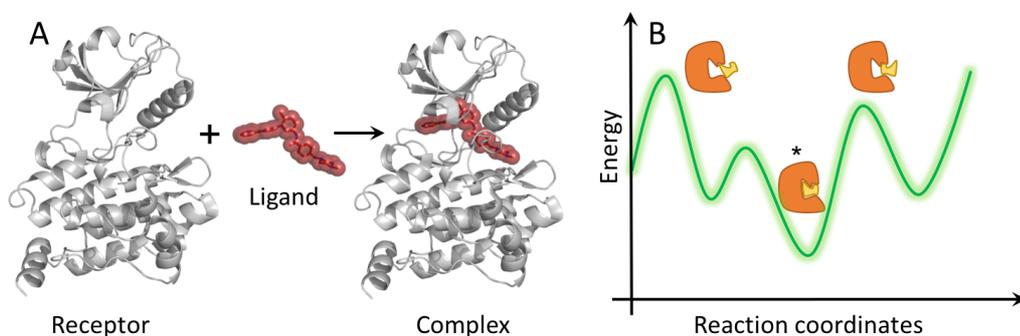


Figure 2.1. A) Representation of complex formation starting from unbound receptor and ligand. B) Schematic representation of the results obtained after posing step of molecular docking simulation. Different ligand-receptor conformations are obtained

having a specific value of energy. Only the conformations within the local minima () are retained for the subsequent scoring phase.*

Posing. During a docking calculation, all possible orientations and conformations of both protein and ligand should be taken into account in order to exhaustively sample the conformational space and retrieve all possible binding modes. In this regard, all six degrees of translational and rotational freedom of the ligand would be explored along with the internal conformational degrees of freedom of both the ligand and protein. However, this is not feasible with the computational resources available nowadays. In fact, for a ligand with four rotatable bonds and six rigid-body alignment parameters, it has been estimated that approximately 2000000 years of computational time (at a rate of 10 configurations per second) are required to make an exhaustive conformational search.⁸¹ To reduce the overall computational cost, two strategies are employed. The first one is to treat only the ligand as flexible and consider the receptor as rigid or semi-rigid body (allowing for example the movement of selected side chains within the receptor). The second strategy is to employ algorithms to reduce computational cost of the conformational sampling step.⁸² In this regard, several algorithms are currently used, where the most common are: metropolis Monte Carlo (MC) methods, genetic algorithms (GA) and fragment-based methods. MC methods apply random Cartesian moves to the system, accepting or rejecting the move based on Boltzmann probability. ICM⁸³ and Glide⁸⁴ are examples of software that use MC methods. In the GA algorithm, the evolution of a population of possible solutions via genetic operators (mutations crossovers and migrations) to a final population is predicted optimizing a predefined fitness function. This method is used in GOLD software.⁸⁵ Lastly, in fragment-based methods a molecule is divided in portions, which are docked and subsequently linked to reconstruct the entire molecule. FlexX⁸⁶ and DOCK⁸⁷ are software that employ this method.⁸¹ As final result of the conformational search procedure, a set of poses that represent the local minima are returned and subsequently submitted to the scoring phase (Figure 2.1 B).

Scoring. The aim of the scoring function is to distinguish between the correct and incorrect ligand poses, as well as between the active and not active molecules in a reasonable computation time. To do so, the scoring functions estimate the binding

affinity between the ligand and the receptor (ΔG_{bind}), adopting various assumptions and simplifications. Scoring functions can be divided in force-field-based, empirical and knowledge-based.⁸⁰

Classical force-field-based scoring functions estimate the binding energy by computing the sum of non-bonded interactions (electrostatics and van der Waals).⁸⁸ Electrostatic interactions are calculated through Coulombic formalism, while van der Waals are described through Lennard-Jones potential function. Extension of force-field-based scoring functions also take into account of hydrogen bonds, solvations and entropy contributions. Software as DOCK, GOLD and AutoDock employ these scoring functions.

The empirical scoring functions compute the binding energy as a sum of different terms as hydrogen bond, ionic interaction, hydrophobic effect and binding entropy.⁸⁹ The individual terms are rescaled through empirical coefficients obtained by regression analysis from dataset of known ligand-receptor complexes. ChemScore, PLP and LUDI are examples of empirical scoring functions.

Lastly, the knowledge-based scoring functions use released crystal structures complexes to obtain statistical meaningful information on interatomic contact frequencies and/or distances between the ligand and protein.⁹⁰ In practice, they are based on the assumption that the more favorable an interaction is, the greater the frequency of occurrence will be. These frequency distributions are further converted into pairwise atom-type potentials. PMF, DrugScore and SMOG are an example of knowledge-based scoring functions.

However, typical scoring functions do not treat adequately the solvation effect, negatively influencing the affinity prediction. In this case, a rescoring of poses using a physics-based scoring as MM-PB/SA and MM-GB/SA (MM stands for molecular mechanics, PB and GB for Poisson-Boltzmann and Generalized Born, respectively, SA for solvent-accessible surface area), could improve the accuracy of binding affinity prediction.⁹¹

Molecular docking is an essential tool for the structure-based drug design (SBDD), being the fundamental technique employed for Virtual Screening (VS) campaigns. In VS, a library of small molecules is docked into the target receptor in order to identify hit compounds (i.e. a set of compounds that most likely make favorable interactions with the target protein).⁹² These compounds are subsequently synthesized or

purchased and then submitted to biological assay. Compared to the classical high-throughput screening (HTS), where a set of compounds are directly tested *in vitro*, the VS has a number of advantages because is less time demanding, is low cost and allows the evaluation of a broader chemical space.

2. Molecular Dynamics simulations

Molecular dynamics (MD) simulations are computational tools employed to study the physical basis of the structure and the function of biological systems.²⁹ In other words, MD is used to obtain important information on the time evolution of conformations of biological macromolecules as well as kinetic and thermodynamic insights.⁹³ To do so, MD simulates the atomic motions, exploiting the equation of motion defined in classical (i.e. Newtonian) mechanics:

$$F_i = m_i a_i \quad (1)$$

Where F_i is the force applied on particle i , m_i is the mass of particle i , while a_i is the acceleration of particle i . Since the force can be also expressed as the gradient of the potential energy, the previous formula can be written as:

$$-\frac{dU}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \quad (2)$$

Where U is the potential energy of the system. MD simulations can be seen as an iterative process which consists of three stages:

1) A model of the molecular system is prepared based on X-ray, NMR or homology-modeling data. To this, initial random velocities, which depend on the temperature of the system, are assigned.

2) The forces acting on each atom are computed from an equation like:

$$\begin{aligned}
U = & \sum_{bonds} k_r (r - r_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 \\
& + \sum_{dihedrals} k_\phi [1 + \cos(n\phi + \phi_0)] \\
& + \sum_{atom\ i} \sum_{j \neq i} 4\varepsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^6 \right] + \sum_i \sum_{j \neq i} 4\varepsilon_{i,j} \frac{q_i q_j}{\varepsilon_0 r_{i,j}}
\end{aligned} \tag{3}$$

In brief, the total energy is a contribution of bonded and non-bonded interactions. For computational feasibility, the potential is taken to be pair-wise additive. Chemical bonds and atomic angles are modeled using simple virtual springs, while dihedral angles are modeled using a sinusoidal function. Non-bonded forces are composed by van der Waals interactions, modeled using the Lennard-Jones 6-12 potential⁹⁴ and electrostatic interactions, which are modeled using Coulomb's law. In order to reproduce the real behavior of molecules in motions, the energy terms described above are parameterized to fit quantum-mechanical calculations and experimental data. Together, these parameters are called 'force fields'. To date, several force fields are routinely used in MD simulations, including AMBER,⁹⁵ CHARMM⁹⁶ and GROMOS.⁹⁷

3) The Newton's equation of motion (1) is solved and the new atomic positions and velocities are updated.

Stages 2) and 3) are iterated several times (i.e. steps or Δt) in order to have the molecular trajectory.⁹⁸ MD is a deterministic method, inasmuch the state of the system at any future time can be predicted from its current state.

From MD simulations, the time average properties A of a system are calculated. In accordance with the ergodic hypothesis, which is one of the fundamental axioms of statistical mechanics, the ensemble average of a property A is equal to the time average (4).

$$A_{avg} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t dt' A(r; t') \tag{4}$$

MD simulations present two main issues: the first one is related to the approximations of the current force field, while the second is the computational cost associated to simulate biological relevant time scale. In this respect, many important

biological events, as protein folding, drug binding, protein conformational changes, take place within timescales much longer than the time length routinely simulated. However, the use of novel hardware as well as graphical processing units (GPUs) enables to overcome the time-scale limitations giving the possibility of run microsecond-long MD simulations. In this regard, the group of DE Shaw is one of the pioneers of this approach. In fact, thanks to their supercomputer Anton they are able to simulate millisecond of simulations, capturing events as protein folding/unfolding⁹⁹ and drug binding.¹⁰⁰

Besides, enhanced sampling techniques are widely used methods that enable to explore the protein conformational landscape using nanosecond or microsecond long MD simulations. The main idea behind these methods is the application of an external bias to the potential energy of the system, which allows lowering the barrier energies between different conformational states and exploring different protein configurations in a reasonable time. Examples of these methods are metadynamics and accelerated MD simulations, both employed in this thesis.

2.1. Metadynamics

In metadynamics simulations an additional bias potential (or force) is added on a selected number of degrees of freedom, referred to as collective variables (CVs), in order to escape from energetic minima and to facilitate the sampling of the free energy landscape. The external potential bias is history dependent and can be written as a sum of Gaussians deposited in the CV space to discourage the system from revisiting configurations that have already been sampled. The algorithm can be easily understood using Laio's metaphor of the walker, which falls in a swimming pool in the night.¹⁰¹ Here, the walker has access to a large amount of sand (i.e. Gaussians) that can use to fill the swimming pool. When the swimming pool is full of sand, finally the walker is able to climb out of it.

Let S be a set of d functions of the microscopic coordinates R of the system (5):

$$S(R) = (S_1(R), \dots, S_d(R)) \quad (5)$$

At time t , the external potential can be written as:

$$V_G(S, t) = \int_0^t dt' \omega \exp\left(\frac{\left(S_i(R) - S_i(R(t'))\right)^2}{2 \sigma_i^2}\right) \quad (6)$$

Where σ_i is the width of the Gaussian for the i th CV, while ω is a constant energy rate expressed as the ratio between the Gaussian height and the deposition stride τ_G :

$$\omega = \frac{W}{\tau_G} \quad (7)$$

The Figure 2.2 shows a schematization of the metadynamics technique. The walker (green dot) explores the unknown free energy landscape and Gaussians are added to allow overcoming high energy barriers (a). Once that the free energy surface (FES) is completely filled (b), the FES is reconstructed by summing the deposited Gaussians (c). (c).

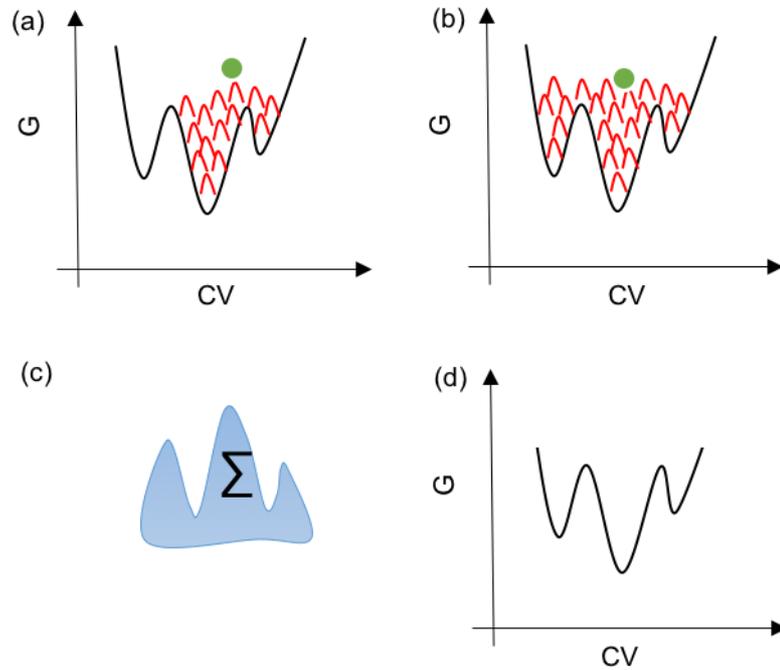


Figure 2.2. Schematic representation of the metadynamics technique. Figure adapted from Van Speybroeck et al.¹⁰²

In fact, when the metadynamics reaches the convergence (i.e. the free energy landscape is filled), the V_G provides an unbiased estimate of the underlying free energy:

$$\lim_{t \rightarrow \infty} V_G(S, t) \sim -F(S) \quad (8)$$

The free energy $F(S)$ is defined as:

$$F(S) = -\frac{1}{\beta} \ln \left(\int dR \delta(S - S(R)) e^{-\beta U(R)} \right) \quad (9)$$

Where $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant, T is the temperature of the system and $U(R)$ is the potential energy function.

However, metadynamics simulations have two main drawbacks. The first one is related to the correct choice of the CVs. Indeed, to guarantee an effective application of metadynamics, the CVs must respect the following guidelines:

- They should distinguish between the initial and final state and describe all the relevant intermediates.
- They should describe all the slow events that are relevant to the process of interest.
- They should be limited in number, otherwise it will take a very long time to fill the FES.

In this respect, the choice of a wrong CVs or neglecting an important CV could increase enormously the time required to reach the convergence to the FES and could also lead to a typical hysteric behaviour of the reconstructed FES.¹⁰³ Collective variables can span from the simplest as distances, angles and radius of gyration to most complex as path variables and contact map.

Another weakness of metadynamics is that it is difficult to decide when to terminate a metadynamics run. In fact, in a single run, the V_G does not converge to a definite value but fluctuates around the correct result, leading to an average error which is proportional to the square root of the bias potential deposition rate.^{104, 105} In addition, extending too much a metadynamics simulation could irreversibly push the system in regions of configurational space which are not physically relevant. Interestingly, the well-tempered metadynamics method offers a solution for this

issue. In fact, thanks to this approach, it is possible to obtain an estimation of the FES that converges to the exact result in the long time limit.¹⁰⁶ In detail, in well-tempered metadynamics the bias deposition rate decreases over simulation time and the external bias assumes the form:

$$V(S, t) = \Delta T \ln \left(1 + \frac{\omega N(S, t)}{\Delta T} \right) \quad (10)$$

where ω has the dimension of an energy rate, ΔT an input parameter with the dimension of a temperature and $N(S, t)$ is the histogram of the S variables collected during the simulation. In practice, the Gaussian height W is scaled according to:

$$W = \omega \tau_G e^{-\frac{V_G(S, t)}{k_B \Delta T}} \quad (11)$$

Through well-tempered metadynamics, the bias potential does not fully compensate the FES, but it converges to:

$$V_G(S, t \rightarrow \infty) = -\frac{\Delta T}{T + \Delta T} F(S) + C \quad (12)$$

where C is an immaterial constant. Therefore for $\Delta T \rightarrow 0$, ordinary MD is recovered, whereas the $\Delta T \rightarrow \infty$ limit corresponds to standard metadynamics. By tuning the ΔT in between 0 and ∞ one can regulate the extent of FES exploration, avoiding the overfilling.

2.2. Accelerated MD

In accelerated MD (aMD) simulations a continuous non-negative bias boost potential function $\Delta V(r)$ is added to the true potential energy $V(r)$, in order to accelerate the dynamics of the system and favor the transition of high energy barriers (13). Therefore, the resulting potential energy $V^*(r)$ is given by the sum of the true potential energy $V(r)$ and the external bias $\Delta V(r)$. The external bias is applied only

when $V(r)$ is below a certain threshold E (see Figure 2.3 A), according to the following scheme:

$$V^*(r) = \begin{cases} V(r), & V(r) \geq E \\ V(r) + \Delta V(r), & V(r) < E \end{cases} \quad (13)$$

Hence, when the true potential energy $V(r)$ of the system is higher or equal than the reference energy E no bias is added and the simulation proceeds as a classical MD scheme (13). On the contrary, when the true potential energy is smaller than the reference energy, the bias boost is added in order to accelerate the system. The bias boost has the following form:

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + E - V(r)} \quad (14)$$

where the term α is tuning parameter. The presence of the parameter α enables to fill the energetic minima, maintaining the underlying shape of the unmodified potential energy surface. In detail, when α has a high value, the bias boost smoothly merges with the unmodified potential energy surface. Instead, when α is zero the modified potential energy $V^*(r)$ is equal to E , which corresponds to a solution proposed by Rahman and Trully in 2002 (see Figure 2.3 B).¹⁰⁷

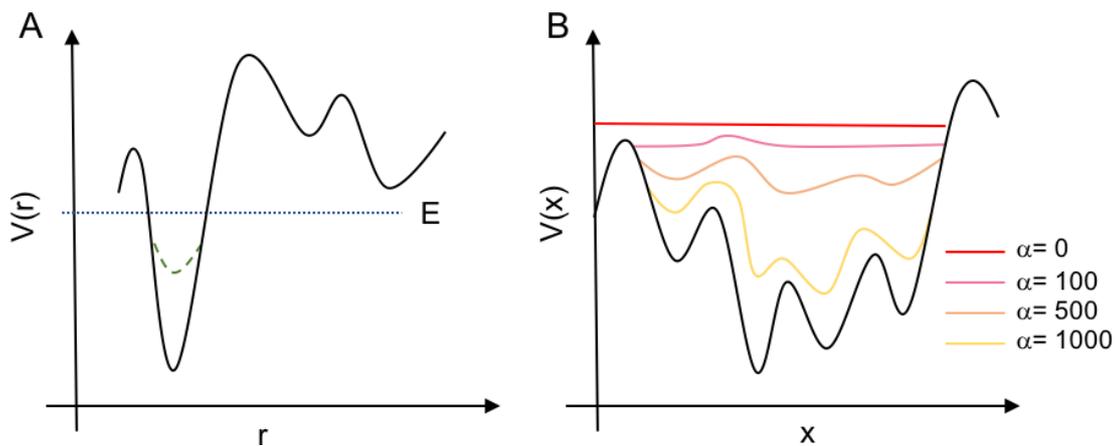


Figure 2.3. A. Representation of the true potential $V(r)$ (black line), of the threshold E (dotted line) and of the bias $\Delta V(r)$ (dashed line). B. Representation of potential energy function (black line) and several bias (dotted and dashed lines) plotted at different values of α . Figure adapted from Hamelberg D. et al.¹⁰⁸

In the first application of the aMD simulation, carried out by the McCammon group, the external bias potential energy was applied only to the dihedral term of the potential energy function, according the rationale that the conformational changes of a biomolecular systems are determined chiefly from changes in the torsional.¹⁰⁸ However, this method has a strong limitation when the biomolecule is surrounded by solvent. In fact, the large conformational changes of the biomolecule are slowed by the diffusion of the solvent around the solute. To overcome this issue, in 2007, the same group developed a variant of the method in which the bias was applied to the whole potential. Thanks to this modification, the authors were able to sample accurately diffusion of the solvent.¹⁰⁹

The choice of both E and α has a great importance, inasmuch they determine the acceleration of the simulation. In principle, E has to be higher than minimum potential energy (V_{\min}), otherwise a classical MD simulation is performed. To retrieve V_{\min} , a short classical MD is performed and the resulting average potential energy is considered as V_{\min} .¹⁰⁸ Although several trials are required to choose the optimal values of both E and α , a general formula has been developed with the aim of retrieving values as close as possible to the optimal ones:

$$E_{dihе} = V_{dihе_avg} + 4 \times N_{res} \quad (15)$$

$$\alpha_{dihе} = 4 \times \frac{N_{res}}{5} \quad (16)$$

$$E_{total} = V_{total_avg} + 0.16 \times N_{atoms} \quad (17)$$

$$\alpha_{total} = 0.16 \times N_{atoms} \quad (18)$$

where N_{res} is the number of residues of the solute, N_{atoms} is the total number of the atoms, $V_{dihе_avg}$ and V_{total_avg} are the average of dihedral and total potential energy.

Finally, aMD is also able to return the canonical average of an observable $A(r)$. Using the modified potential $V^*(r)$, the ensemble average of the observable $A(r)$ is:

$$\langle A^* \rangle = \frac{\int dr A(r) e^{-\beta V^*(r)}}{\int dr e^{-\beta V^*(r)}} \quad (19)$$

by substituting $V^*(r)$ in (19), it becomes:

$$\langle A^* \rangle = \frac{\int dr A(r) e^{-\beta V(r) - \beta \Delta V(r)}}{\int dr e^{-\beta V(r) - \beta \Delta V(r)}} \quad (20)$$

To obtain the corrected ensemble average, the phase space of the modified potential is reweighted by multiplying each configuration by the strength of the bias at each position.

$$\langle A^C \rangle = \frac{\int dr A(r) e^{-\beta V(r) - \beta \Delta V(r)} e^{\beta \Delta(r)}}{\int dr e^{-\beta V(r) - \beta \Delta V(r)} e^{\beta \Delta(r)}} \quad (21)$$

$$\langle A^C \rangle = \frac{\int dr A(r) e^{-\beta V(r)}}{\int dr e^{-\beta V(r)}} = \langle A \rangle \quad (22)$$

Although the aMD is an excellent method to enhance the sampling of the phase space of a biomolecular system, the reweighting procedure suffers several limitations making difficult to recover the original free energy landscape, specially for large systems. This is due to the fact that in aMD the bias energy is typically large, reaching also hundreds of calories per mole, leading to noise during the reweighting procedure. To overcome this issue, in 2015 Miao et al. developed a variant of aMD, called Gaussian accelerated MD (GaMD), which enables to reduce the energetic noise. In brief, according to this new method the bias boost assumes Gaussian distribution that allows a more accurate reweighting respect to the classic aMD.¹¹⁰

Chapter 3

HRD motif as the central hub of the signaling network for activation loop autophosphorylation in Abl kinase

ABSTRACT

A number of structural factors modulate the activity of Abelson (Abl) tyrosine kinase, whose deregulation is often related to oncogenic processes. First, only the open conformation of the Abl kinase domain's activation loop (A-loop) favors ATP binding to the catalytic cleft. In this regard, the *trans*-autophosphorylation of the Y412 residue, which is located along the A-loop, favors the stability of the open conformation, in turn enhancing Abl activity. Another key factor for full Abl activity is the formation of *active* conformations of the catalytic DFG motif in the Abl kinase domain. Furthermore, binding of the SH2 domain to the N-lobe of the Abl kinase domain was recently demonstrated to have a long-range allosteric effect on the stabilization of the A-loop open state. Intriguingly, these distinct structural factors imply a complex signal transmission network for controlling the A-loop's flexibility and conformational preference for optimal Abl function. However, the exact dynamical features of this signal transmission network structure remain unclear. Here, we report on microsecond-long molecular dynamics coupled with enhanced sampling simulations of multiple Abl model systems, in the presence or absence of the SH2 domain, and with the DFG motif flipped in two ways (*in* or *out* conformation). Through comparative analysis, our simulations augment the interpretation of the existing Abl experimental data, revealing a dynamical network of interactions that interconnect SH2 domain-binding with A-loop plasticity and Y412 autophosphorylation in Abl. This signaling network engages the DFG motif and, importantly, other conserved structural elements of the kinase domain, namely the EPK-ELK H-bond network and the HRD motif. Our results show that the signal propagation for modulating the A-loop spatial localization is highly dependent on the HRD motif conformation, which thus acts as the central hub of this (allosteric) signaling network controlling Abl activation and function.

INTRODUCTION

The Abelson (Abl) tyrosine kinase is a tightly regulated nonreceptor kinase involved in a variety of cellular pathways that control cell growth, survival, and morphogenesis.^{111, 112} The fusion of Abl and the breakpoint cluster region (Bcr) genes leads to expression of Bcr-Abl, a hyperactivated protein kinase that lacks certain auto-inhibitory mechanisms and is the main driver of chronic myelogenous leukemia (CML). For this reason, the main therapeutic avenue for treating CML uses tyrosine kinase inhibitors, such as imatinib, to inhibit Bcr-Abl kinase activity.^{113, 114, 66, 115}

The Abl kinase has a modular structure comprising the two regulatory SH3 and SH2 domains and a catalytic kinase domain (KD).⁶⁹ Like all kinases, KD in Abl is formed by the N- and C-lobes, which jointly form a deep cleft where ATP binds. Then, the phosphotransfer reaction occurs, with the γ -phosphate of the ATP transferred onto a tyrosine residue of the substrate proteins, which regulates the Abl signaling pathways.^{67, 68, 116, 117} The activity of Abl is highly controlled by extensive allosteric networks¹¹⁸ which, in turn, can be affected by drug binding¹¹⁹, as well as point mutations. Importantly, the catalytic process is highly dependent on the conformation of the flexible activation loop (A-loop). In Abl, this is a sequence of 29 residues bounded by the conserved D400-F401-G402 (DFG) and A426-P427-E428 (APE) motifs. ATP and the peptide substrate can only bind and undergo the phosphotransfer reaction when the A-loop is in its fully open conformation. In this regard, a key chemical step for Abl activation is the trans-autophosphorylation at the Y412 residue located along the A-loop. Several experimental and computational studies have in fact demonstrated that phosphorylation at the Y412 stabilizes the A-loop in a fully open conformation, promoting ATP binding and, thus, full kinase activation.^{120, 121} Indeed, trans-autophosphorylation at the Y412 has been found to increase the Abl catalytic activity by ~ 10 -fold.^{122, 123, 124} Moreover, Abl function also depends on the conformation adopted by the DFG motif in KD. Only the DFG-*in* conformation, in which the D400 coordinates the catalytic magnesium ion, allows the phosphotransfer reaction.^{125, 57, 58} In this conformation, the F401 side chain faces the ATP binding site, allowing the proper formation of the regulatory spine, i.e. R-spine (a conserved motif which in Abl is formed by M309, L320, F401, H380 and D440) whose proper assembly is known to be important for kinases catalytic function.⁵⁸ Conversely, in the DFG-*out* conformation, the flipped D400 side chain

causes the loss of catalytic kinase function and, also, the reorientation of F401 side chain leads to the disengagement of R-spine (Figure 3.1).

Finally, a *strained* conformation of the well-conserved H380-R381-D382 (HRD) motif, located close to the ATP binding site, was recently reported to be crucial for full Abl activity. In this conformation, the HRD backbone interacts with the D440 of the α F helix via a conserved H-bond network, called EPK-ELK^{78, 72} (Figure 3.1). The presence of this network allows the R381 side chain to interact via a salt bridge with the phosphorylated Y412, stabilizing the fully open conformation of the A-loop, which in turn favors Abl activity.¹²⁶

In addition to these key structural and catalytic KD elements, the SH2 domain has also been found to significantly affect Abl function, thereby acting as a modulatory subunit.^{76, 127} Indeed, structural and biochemical investigations have shown that the SH2 binding to the N-lobe of KD in Abl^{72, 70, 119} (Figure 3.1) favors the A-loop phosphorylation, increasing the kinase catalytic activity. The SH2 domain affects the A-loop motility stabilizing the fully open conformation, which is thought to be optimal for the trans-autophosphorylation reaction.¹²⁸ This was shown by Dölker *et al.*¹²⁹ who used MD simulations integrated with mutagenesis experiments to demonstrate that SH2 affects the motion of several catalytically important portions of KD, which in turn favor catalytically-active conformations of Abl. The β 3- α C loop was identified as a key structural motif in transmitting the allosteric signal from the SH2 domain to the KD. More recently, Tse *et al.*¹³⁰ employed molecular simulations and the network-based modeling to describe the subtle thermodynamic balance of conformational states that shape the allosteric signal within KD, favoring Abl activity. While these informative studies describe some of the allosteric effects of SH2 domain on Abl activity, how exactly key and distal structural motifs are dynamically interconnected so to modulate and favor the A-loop trans-autophosphorylation in Abl remains poorly defined. What affect A-loop spatial conformation and Y412 propensity for phosphorylation? What are the possible molecular mechanisms that form and modulate a signal transmission network structure for the A-loop spatial orientation and Abl activity?

To further address these questions, we coupled microsecond-long classical molecular dynamics (MD) with free-energy simulations to decipher a structured and dynamic allosteric network of interactions, which appears to be critical for

modulating A-loop plasticity, in turn regulating the phosphorylation at Y412.^{131, 132} This complex network of dynamic interactions was revealed through a comparative analysis performed using several different systems of Abl, alone or in complex with the modulatory SH2 domain, and with the DFG-*in* or DFG-*out* conformations of KD. This signaling network engages the DFG motif and, importantly, other conserved structural KD elements, such as the EPK-ELK H-bond network and the HRD motif. Overall, our simulations suggest that the HRD motif acts as the central hub of the allosteric signaling network in Abl, acting cooperatively with the DFG motif and the EPK-ELK H-bond network to regulate A-loop accessibility.

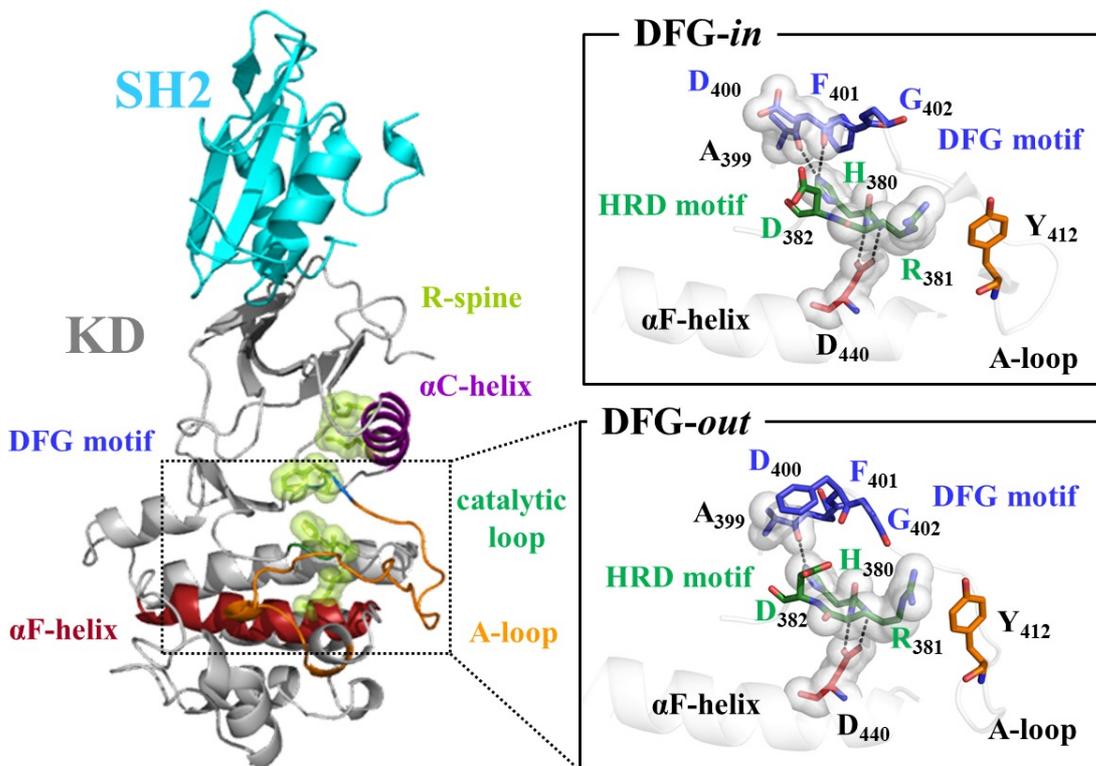


Figure 3.1. Left: overview of the Abl kinase domain (KD) (grey) bound to the SH2 domain (cyan). PDB code: 1OPL chain B. Colors of KD indicate: α C helix (violet), DFG motif (blue), catalytic loop (green), A-loop (orange), α F helix (red), R-spine (lime). Right: a close view of the network formed by DFG, HRD, and α F helix, both in DFG-*in* (upper box, PDB code: 2F4J) and DFG-*out* (bottom box, PDB code: 1OPL chain B) conformations.

RESULTS

Molecular dynamics (MD) simulations. We performed a comparative MD-based analysis of three different Abl systems: the SH2/KD complex in the catalytically inactive DFG-*out* conformation (SH2-KD_{out} system), the KD alone in the DFG-*out* (KD_{out} system), and the KD alone in the DFG-*in* conformation (KD_{in} system, see Methods for further details). We wished to clarify how the regulatory subunit SH2's presence or absence and the DFG motif's initial conformation (*in* or *out*) affect A-loop spatial conformation and Y412 propensity for phosphorylation, which is needed for full Abl activation.^{122, 123, 124} The Y412 conformational states are here considered to be prone to the trans-autophosphorylation according to two key parameters: 1) the RMSD of Y412 with respect to the reference X-ray structure 1OPL chain B; 2) the distance between the Y412 and D440 along the α F helix (see Methods for further details).

SH2-KD_{out} dynamics. Extensive MD simulations of over $\sim 1.8 \mu\text{s}$ of the SH2-KD_{out} system highlighted the elevated stability of this complex (see RMSD in Figure 3.2 and Figure S1). We used a principal component analysis (PCA) to examine the dominant motions of the system during the MD simulations. In agreement with previous studies, the first eigenvectors describes the so-called “breathing” (or “hinge”) and the “lobe-twisting” motions, which are well known to be essential for the kinase activity (Figure 3.3).^{129, 133, 134, 55, 135, 136, 137, 138}

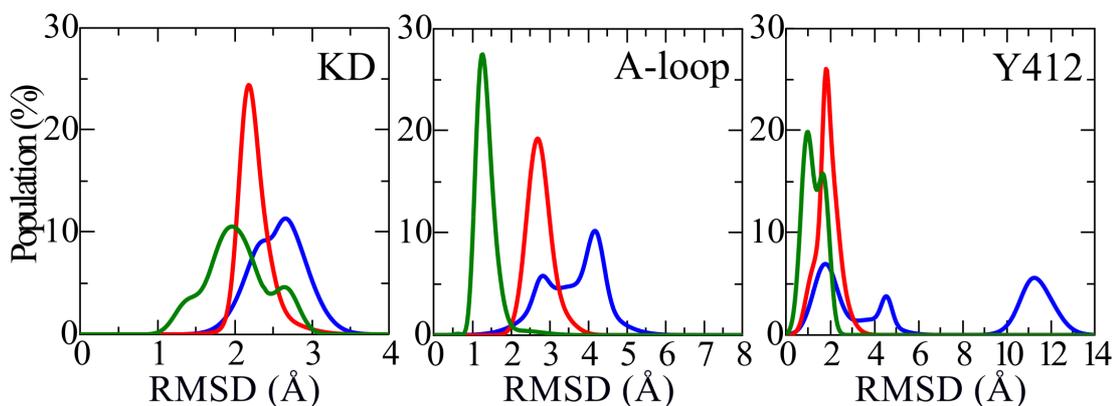


Figure 3.2 Comparison of RMSD frequency distribution of the three systems (SH2-KD_{out}, red; KD_{out}, blue; KD_{in}, green) for the KD backbone atoms (left panel), A-loop backbone atoms (central panel), and Y412 (right panel).

Throughout the simulation, the A-loop remained in its starting fully open conformational state (see RMSD in Figure 3.2 and Figure S1). As a consequence, Y412 was also stable, as indicated by the single deep basin obtained by plotting the $d[Y_{412}-D_{440}]$ values against $RMSD_{Y412}$ (Figure 3.4). In detail, $RMSD_{Y412}$ was $1.89 \pm 0.51 \text{ \AA}$ (Figure 3.2 and Figure S1), while $d[Y_{412}-D_{440}]$ oscillated around $12.97 \pm 0.95 \text{ \AA}$ (Figure 3.5), vs. 12.60 \AA in the X-ray 1OPL chain B. Hereafter, this conformation is referred to as *Conf-1*. In addition, the H-bond network, formed by the two antiparallel β sheets $\beta 10$ and βA -loop (formed when the A-loop is in the fully open conformation), was preserved for throughout the MD trajectory (Figure S2), underlining this conformation's stability.

Another structural feature that was maintained in its crystallographic position throughout the MD simulations was the backbone geometry of the HRD motif (see RMSD in Figure S3). This motif is known to be important for proper kinase catalytic activity.¹²⁶ In particular, the backbone torsional angle ϕ of R381 (i.e. $R381\phi$) is stable, being equal to $59.01^\circ \pm 8.02^\circ$ (vs. 73.8° in the X-ray 1OPL, see Figure 3.5). With this HRD conformation, hereafter referred to as *native*, the R381 side chain points toward Y412. This was reflected in the $d[Y_{412}-R_{381}]$ value, i.e. the distance between the phenolic oxygen of Y412 and the guanidinium carbon of R381 (Figure S4), which was stably maintained at $5.61 \pm 1.21 \text{ \AA}$ (vs. 3.90 \AA in the X-ray 1OPL). The EPK-ELK H-bond network directly connects the HRD backbone with the αF helix (see Methods for details). This network was also preserved during the MD simulations. The two H-bonds that form the EPK-ELK network, namely $Hb[H_{380}-D_{440}]$ and $Hb[R_{381}-D_{440}]$, were steady at $3.46 \pm 0.20 \text{ \AA}$ and $2.98 \pm 0.16 \text{ \AA}$, helping to maintain the overall stability of the *native* HRD conformation (Figure 3.5). Finally, $Hb[H_{380}-A_{399}]$ is an H-bond between the $\epsilon 2$ nitrogen of side chain of H380 and the carbonyl backbone of the DFG-1 residue A399 (Figure 3.1). $Hb[H_{380}-A_{399}]$ was firmly at $3.32 \pm 0.64 \text{ \AA}$ (Figure 3.5) reflecting, in turn, an overall stability of the H380 side chain conformation.¹³⁹

Overall, these results demonstrate the stability of the SH2-KD complex, with full preservation of both the crystallographic open A-loop conformation, in which the Y412 is prone to be trans-autophosphorylated, and the *native* HRD conformation, even in the absence of the dasatinib-like inhibitor (see Methods).

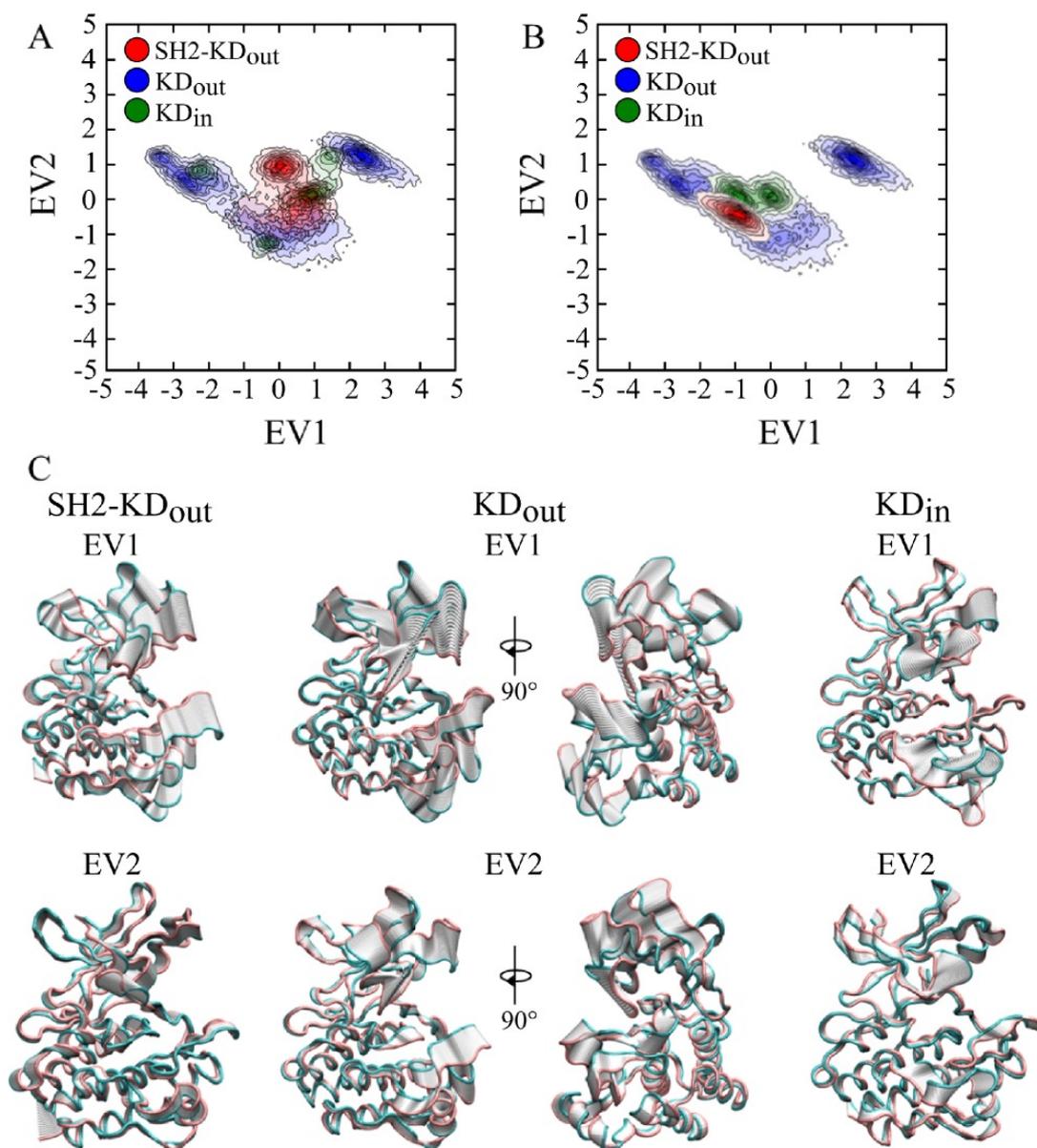


Figure 3.3. A) Projection of MD trajectories of SH2-KD_{out} (red), KD_{out} (blue), and KD_{in} (green) along the first two eigenvectors of PCA analysis. B) Projection of the same MD trajectories along the first two eigenvectors of KD_{out}. C) The principal motion projected along both EV1 and EV2 is superimposed sequentially with 20 frames of the whole trajectory of SH2-KD_{out}, KD_{out}, and KD_{in}, respectively. The dominant movements are the so-called “breathing” and “lobe-twisting” motions. The former describes the opposite movements of the two KD lobes, with alternate closure and opening of the catalytic cleft, while the latter evidences the N- and the C-lobe twist against each other.

KD_{out} dynamics. After removing the SH2 regulatory subunit, the A-loop displayed an augmented motility (see RMSD Figure 3.2 and Figure S1). The projection of the MD trajectory onto the first two eigenvectors shows an extensive sampling of phase space, with the system visiting different minima (Figure 3.3). Here, EV1 and EV2 describe more pronounced breathing and twisting motions compared to that observed in SH2-KD_{out} (Figure 3.3). They also suggest that the A-loop visited a larger conformational space (Figure 3.3), indicating its overall augmented flexibility compared to SH2-KD_{out} system. Indeed, in KD_{out}, the A-loop explored different spatial orientations. In fact, in addition to the crystallographic fully open conformation, the A-loop also explored a partially closed and a new open intermediate conformation.

In detail, during the first microsecond of simulations, the A-loop is fully open and the Y412 is in an optimal conformation in order to be phosphorylated. This is indicated by $d[Y_{412}-D_{440}]$, which was maintained at $12.24 \pm 1.2 \text{ \AA}$ (Figure 3.5), and by the low RMSD_{Y412} of $1.89 \pm 0.55 \text{ \AA}$ (*Conf-1* in Figure 3.4). Importantly, also the HRD backbone remained in its *native* crystallographic geometry, allowing the R381 side chain to point toward Y412 (Figure S4). Moreover, the EPK-ELK H-bond network was assembled well (Figure 3.5), further stabilizing the *native* HRD geometry.

However, at about $\sim 0.75 \mu\text{s}$, we observed a rotation of the $H380\chi_1$ angle (Figure S5) that caused the definitive rupture of the interaction between the side chain of H380, along the HRD, and the backbone of A399 (DFG-1 residue, Figure 3.5). This event seems to favor a partial closure of the A-loop in which the Y412 is no longer prone to be phosphorylated, because buried in the inner part of KD, pointing toward the D440 of the αF helix (see Figure 3.4). In between ~ 1.00 and $\sim 1.14 \mu\text{s}$, the Y412 assumed the *Conf-2* in which $d[Y_{412}-D_{440}]$ was $5.42 \pm 0.51 \text{ \AA}$ (Figure 3.5) and the RMSD_{Y412} increased to $3.65 \pm 0.57 \text{ \AA}$ (Figure 3.2 and Figure S1). Then, in between ~ 1.14 to $\sim 1.34 \mu\text{s}$, the Y412 directly interacted with the D440 assuming the *Conf-3*, in which $d[Y_{412}-D_{440}]$ shortened to $2.90 \pm 0.26 \text{ \AA}$ (Figure 3.5) and the RMSD_{Y412} was equal to $4.53 \pm 0.25 \text{ \AA}$ (Figure 3.2 and Figure S1). The formation of the *Conf-3* configuration coincided with the rupture of the EPK-ELK H-bond network, as shown by the increased $\text{Hb}[H_{380}-D_{440}]$ and $\text{Hb}[R_{381}-D_{440}]$, which reached $6.00 \pm 0.28 \text{ \AA}$ and $5.36 \pm 0.37 \text{ \AA}$, respectively (Figure 3.5).

The loss of the HRD/DFG-1 interaction, together with rupture of the EPK-ELK H-bond network, generated an overall destabilization of the HRD motif. In fact, the backbone of the HRD motif underwent to a peptide flip, as shown by the rotation of R381 ϕ , which reached a mean value of $-87.66^\circ \pm 33.23^\circ$, then maintained until the end of the simulation (Figure 3.5). The torsion of R381 ϕ generated a reorientation of the R381 side chain, which extended towards the ATP binding site, forming a transient H-bond with the side chain of E305 of the α C helix (Figure S6). Concomitantly with the observed HRD reorientation, the A-loop spontaneously reopened with a major rearrangement that ended in a new intermediate conformation maintained until the end of the simulation ($\sim 2.6 \mu\text{s}$). In this intermediate A-loop conformational state, however, Y412 adopted a different orientation with respect to the starting one, being rotated upside down with respect to the crystallographic orientation and assuming, thus, a non-optimal configuration for the trans-autophosphorylation reaction. Here, $d[\text{Y}_{412}\text{-D}_{440}]$ was equal to $14.66 \pm 1.80 \text{ \AA}$ (Figure 3.5) and the $\text{RMSD}_{\text{Y412}}$ to $11.20 \pm 1.02 \text{ \AA}$ (*Conf-4* in Figure 3.4). Moreover, in this state, the β A-loop was flipped by 180° and shifted by two residues along the β 10 sheet, forming a modified H-bond network that stabilized *Conf-4* (Figure S2). Finally, the EPK-ELK H-bond network was no longer properly assembled, with only $\text{Hb}[\text{H}_{380}\text{-D}_{440}]$ being formed (Figure 3.5). In summary, in the KD_{out} simulations, Y412 was found in a conformation in which is prone to be trans-autophosphorylated for only $\sim 40\%$ of the total simulation time.

To further increase the sampling and validate our findings, we ran two additional MD replicas of $\sim 1.1 \mu\text{s}$ each (Replica 1- KD_{out} and 2- KD_{out}). We began these from a snapshot at $\sim 1.2 \mu\text{s}$ of KD_{out} in which Y412 is in *Conf-3*. This time, the HRD motif did not undergo to a peptide flip, but remained in the *native* conformation (Figure S7C) with the EPK-ELK H-bond network newly restored. Remarkably, the restoring of the *native* HRD conformation allowed the reopening of the A-loop (at ~ 0.30 for Replica 1- KD_{out} and at $\sim 0.55 \mu\text{s}$ for Replica 2- KD_{out}), which reassumed the fully open conformation as in the original X-ray structure (i.e. *Conf-1*, see Figure S7A and B). The A-loop opening restored the crystallographic H-bond networks between the β 10 and β A-loop sheets, which further stabilized the fully open A-loop conformation (Figure S2).

To further evaluate the importance of the EPK-ELK H-bond network in stabilizing the HRD motif, we ran one additional MD simulation of KD, whereby D440, along the α F helix, was mutated into an alanine (herein, D440A-KD_{out} system). This mutation prevented the formation of the EPK-ELK H-bond network, inducing an overall destabilization of the HRD motif (see RMSD in Figure S3). Notably, this time we did not observe any R381 ϕ flip, as in the wild-type KD_{out} simulations (Figure S9C), suggesting that Hb[H₃₈₀-D₄₄₀] formation might be essential for stabilizing the rotated HRD backbone geometry. Thus, the lack of EPK-ELK H-bond network and the subsequent destabilization of the HRD motif seem the major players for triggering a major deformation of the A-loop. Here, indeed, the Y412 adopts a new stable conformation in which is buried in the inner part of KD, assuming a non-optimal configuration for the trans-autophosphorylation (see Figure S8 and S9A for further details).

Finally, we sampled the A-loop conformations through additional MD simulations for over ~ 1.3 μ s, starting from a fully inactive and closed A-loop conformation (herein, closed-KD_{out} system). In this case, the starting conformation corresponded to that crystallized in the presence of imatinib-like inhibitors,^{114, 72, 140, 141, 142, 143} which had not been observed in our aforementioned simulations. In closed-KD_{out}, which is based on the apo form of Abl, we noticed that the closed A-loop conformation was highly unstable (see RMSD in Figure S8). In particular, after ~ 350 ns, the A-loop spontaneously opened, visiting several transient conformations. However, these did not coincide with the fully open state. Along this path, Y412 adopted many different configurations due to the significant flexibility of the A-loop, which mostly assumed conformations that are not prone to be phosphorylated (Figure S9B, and Figure S10).

These results indicate that the absence of the regulatory SH2 subunit greatly increases the overall A-loop flexibility. With DFG-*out*, the A-loop adopts an ensemble of conformations that, even when partially open, are not optimal for trans-autophosphorylation. This explains KD's diminished propensity for autophosphorylation. It also confirms that the inactive and fully closed A-loop conformation is stabilized by the presence of an ATP inhibitor such as imatinib.^{144, 145}

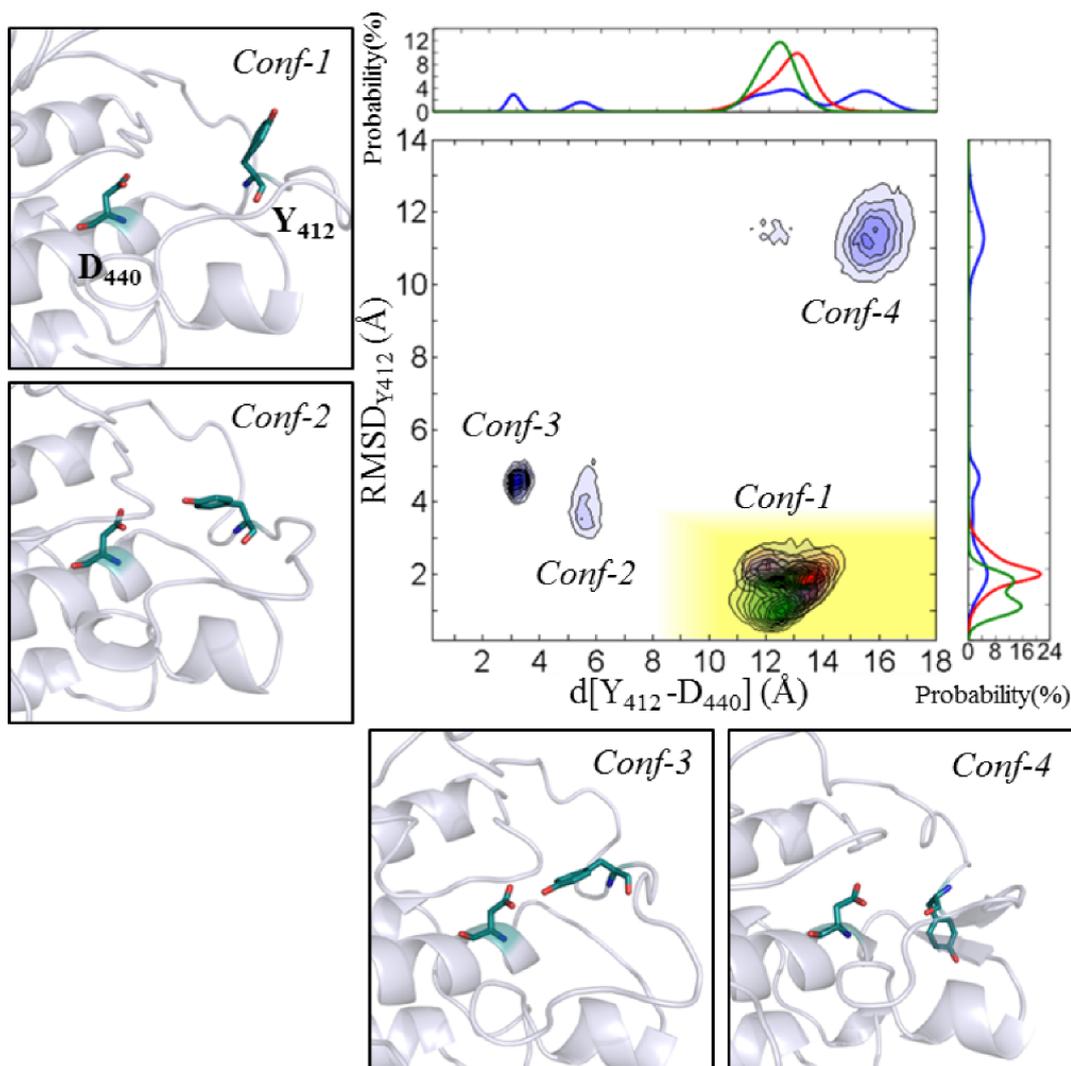


Figure 3.4. 2D plot of SH2-KD_{out} (red), KD_{out} (blue), and KD_{in} (green) MD trajectory data points plotted along $d[Y_{412}-D_{440}]$ –X axes– and the $RMSD_{Y_{412}}$ –Y axes– values. The upper and right panels represent the frequency distribution of $d[Y_{412}-D_{440}]$ and the $RMSD_{Y_{412}}$, respectively. The left and bottom boxes show representative frames of Conf-1, Conf-2, Conf-3, and Conf-4.

KD_{in} dynamics. The overall system was very stable, as shown by the backbone RMSD of KD, with the A-loop in the initial fully open conformation (Figure 3.2 and Figure S1). Here too, the first two eigenvectors described the breathing and lobe-twisting motions (Figure 3.3) similar to those seen for SH2-KD_{out}, which however were less pronounced than KD_{out} (Figure 3.3). Like in SH2-KD_{out}, the H-bond network between the β_{10} and β_A -loop was preserved during all simulations, further suggesting a key contribution in stabilizing the fully open A-loop conformation

(Figure S2). As a result, Y412 was stable in *Conf-1* in which the residue was prone to be phosphorylated (see Figure 3.4).

Additionally, the HRD motif did not undergo substantial fluctuations, preserving its *native* conformation (see RMSD in Figure S3), with the R381 side chain pointing toward Y412. This is indicated by the $R_{381}\phi$ value of $59.01^\circ \pm 8.00^\circ$ and by $d[R_{381}-Y_{412}]$, which was maintained at $5.69 \pm 1.08 \text{ \AA}$ (Figure 3.5 and Figure S4). Similarly, the EPK-ELK H-bond network was well preserved during all simulations, with $Hb[H_{380}-D_{440}]$ and $Hb[R_{381}-D_{440}]$ equal to $3.54 \pm 0.19 \text{ \AA}$ and $2.97 \pm 0.15 \text{ \AA}$, respectively (Figure 3.5). Here, the HRD motif stability is further favored by the stable H-bond between the side chain of the HRD motif's H380 and the carbonyl backbone of the DFG motif's D400, which can be formed only when the DFG is in the *in* conformation. This key interaction, referred to hereafter as $Hb[H_{380}-D_{400}]$, remained stable for all MD simulations, with an average length of $2.94 \pm 0.24 \text{ \AA}$ (Figure 3.5). As a result, KD was maintained in the catalytically fully active conformation for the entire simulation, with Y412 in an optimal state for the trans-autophosphorylation.

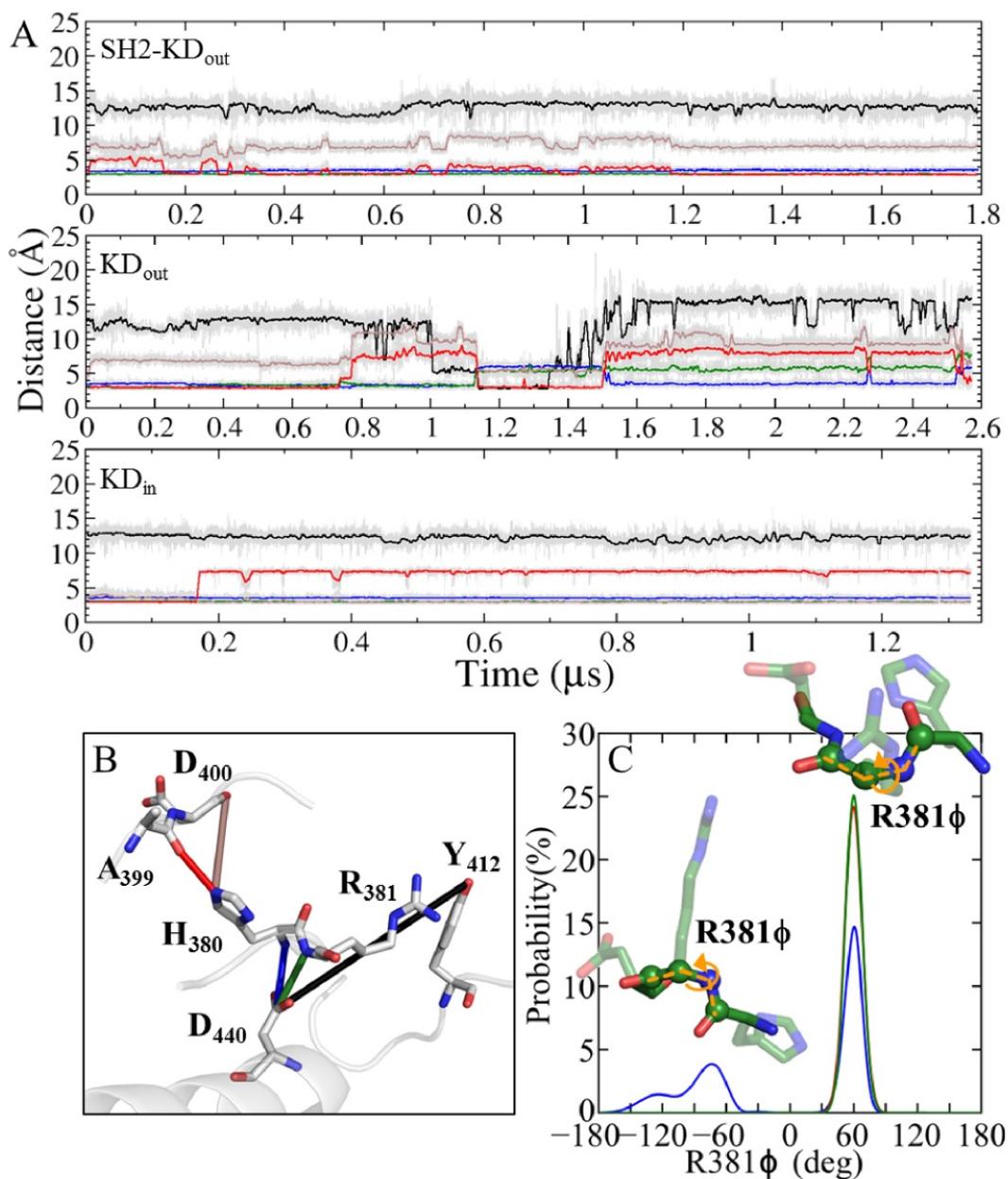


Figure 3.5. A) Distance along time for $d[Y_{412}-D_{440}]$ (black), $Hb[R_{381}-D_{440}]$ (green), $Hb[H_{380}-D_{440}]$ (blue), $Hb[H_{380}-A_{399}]$ (red), $Hb[H_{380}-D_{400}]$ (brown) in the SH2-KD_{out} system (upper panel), KD_{out} system (middle panel), and KD_{in} system (lower panel). B) Graphic representation of the key distances reported in plot A. PDB code: 1OPL chain B. C) Frequency distribution of the R381φ dihedral angle in SH2-KD_{out} (red), KD_{out} (blue), and KD_{in} (green).

Free-Energy Simulations. We used well-tempered metadynamics^{106, 146} to calculate the free-energy surface (FES) related to the A-loop partial closure in the different Abl systems. We employed two walkers for each system (i.e. SH2-KD_{out}, KD_{out} and KD_{in}) and the confined procedure (see Methods and SI) for a total of six independent simulations, starting from either the fully open *Conf-1* or the partially closed *Conf-3* configurations of the A-loop, as observed during our unbiased MD simulations.

In SH2-KD_{out}, four energy minima were detected on the FES. The deepest corresponded to the X-ray conformation in which Y412 adopts *Conf-1* ($d[Y_{412}-D_{440}] \sim 13 \text{ \AA}$), which confirmed this conformation as the most probable and energetically favored one, as observed also in our unbiased MD simulations. Two near minima were found at $d[Y_{412}-D_{440}] \sim 5 \text{ \AA}$ and $\sim 7 \text{ \AA}$, which corresponded to Y412 conformations that are bent toward D440, similarly to *Conf-2*. The last minimum was detected at $d[Y_{412}-D_{440}] \sim 3 \text{ \AA}$, in which Y412 directly interacted with D440, i.e. *Conf-3* (Figure 3.6).

The FES of KD_{out} was characterized by three energetic basins. The global minimum here was *Conf-3*, while *Conf-1* represented a relative minimum on the energy landscape. This was in line with our unbiased MD simulations, which highlighted the stability of *Conf-3*. Another relative minimum was found at $d[Y_{412}-D_{440}] \sim 7 \text{ \AA}$, which represented an ensemble of conformations in which Y412 adopted *Conf-2*. Compared to SH2-KD_{out}, the *Conf-1*→*Conf-2* transition in KD_{out} was energetically more favored, with a slightly lower energy barrier ($\sim 3 \text{ kcal/mol}$ and $\sim 5 \text{ kcal/mol}$ for KD_{out} and SH2-KD_{out} respectively), which might explain why this transition was only observed in our unbiased simulations of KD_{out}, and not in SH2-KD_{out} (Figure 3.6).

In the KD_{in} system, the two deepest minima were found when $d[Y_{412}-D_{440}]$ was between $\sim 9.5 \text{ \AA}$ and 12 \AA . This interval included an ensemble of conformations in which Y412 adopted *Conf-1*. Two other local minima were found at $d[Y_{412}-D_{440}] \sim 5 \text{ \AA}$ and $\sim 3 \text{ \AA}$, which corresponded to *Conf-2* and *Conf-3*, respectively. Here, the *Conf-1*→*Conf-2* transition was energetically more costly than for the SH2-KD_{out} and KD_{out} systems, with a barrier energy of $\sim 8 \text{ kcal/mol}$ (Figure 3.6). This is in line with the high stability of *Conf-1* observed during the unbiased MD simulations.

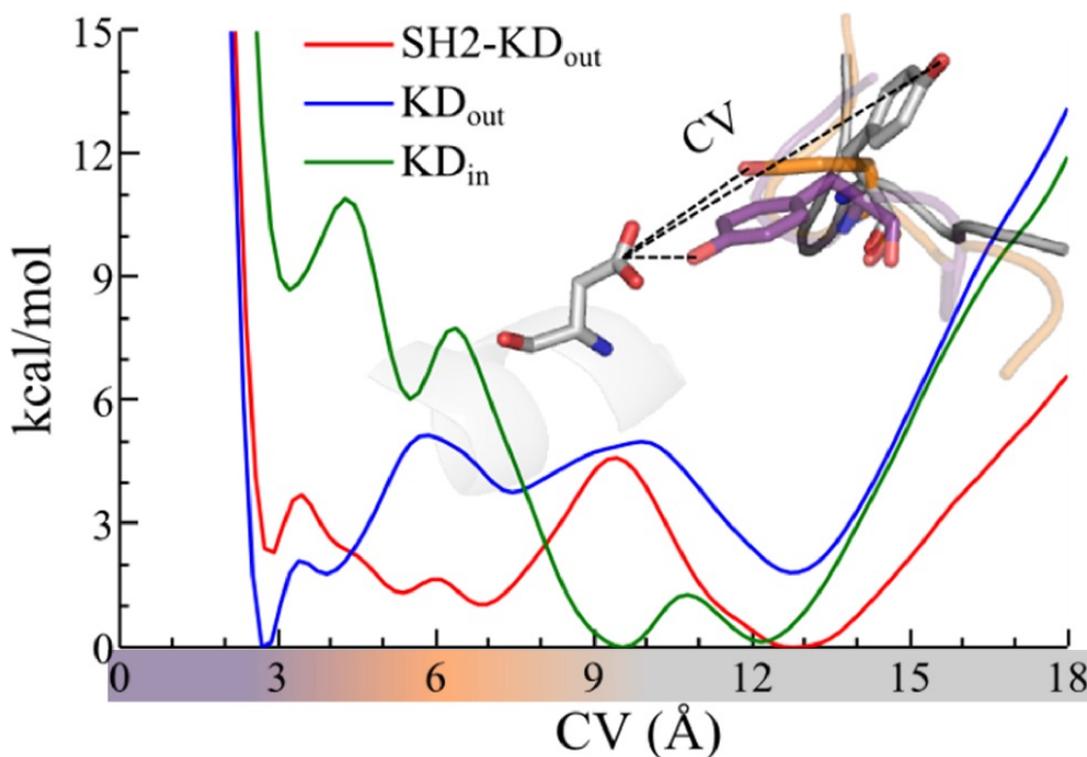


Figure 3.6. Free-Energy Surface (FES) obtained through multiple walkers well-tempered metadynamics simulations for the SH2-KD_{out} (red), KD_{out} (blue), and KD_{in} (green) systems. The CV represents the distance between the center of mass of D440 and the oxygen atom of the Y412 side chain, i.e. $d[Y_{412}-D_{440}]$.

DISCUSSION

In the present work, we performed a comparative analysis using classical molecular dynamics (MD) and free-energy simulations to investigate how the regulatory SH2 domain subunit and the DFG motif (in its *in* or *out* conformations) may affect A-loop spatial arrangements to enable trans-autophosphorylation in Abl, which is necessary for full activation.^{122, 123, 124} We considered the following three different Abl systems (see Methods for further details): i) SH2-KD_{out} complex with KD in DFG-*out* conformation; ii) KD_{out} system with KD alone in DFG-*out* conformation; iii) KD_{in} system with KD alone in DFG-*in* conformation.

The PCA analysis suggested the presence of two large motions in all systems in agreement with previous studies.^{129, 136, 137} These motions are: i) the opening and closing of the two lobes of KD on the active site (i.e. breathing or hinge motion), which is known to be essential for efficient catalysis,^{133, 134, 55} and ii) the twisting of the two lobes of KD against each other (i.e. lobe twisting), which is critical for ADP release (Figure 3.3).^{137, 138} The presence of the SH2 domain in SH2-KD_{out}, and the

preservation of the active DFG-*in* conformation in KD_{in} did not alter these key motions for catalysis. However, in KD_{out}, these two motions are affected and are coupled with the A-loop major conformational flexibility. This likely contributed to the diminished catalytic activity, which has been experimentally observed.^{76, 127}

In fact, in KD_{out}, the A-loop underwent major conformational rearrangements, adopting either fully open (*Conf-1*), partially closed (*Conf-2* and *Conf-3*), or intermediate (*Conf-4*) conformations (see Figure 3.4). Initially, the starting orientation of Y412 in *Conf-1* was gradually flipped by $\sim 90^\circ$ in *Conf-2* and *Conf-3* and, later, by 180° in *Conf-4*, generating a non-optimal conformational state for the Y412 trans-autophosphorylation (see Methods for a definition of Y412 states). Conversely, in both the SH2-KD_{out} and KD_{in} systems, the A-loop remained stable in the fully open *Conf-1* throughout our simulations. Additionally, the computed free-energy landscape indicated that the partial closure of the A-loop (i.e. the *Conf-1*→*Conf-2* transition) was particularly favored in the KD_{out} system. In SH2-KD_{out} and KD_{in}, this transition was energetically costly, in agreement with the unbiased simulations of those systems, which did not show such a transition. Taken together, these results indicate that the fully open A-loop conformation is favored by the presence of the SH2 domain or by the preservation of the active DFG-*in* conformation. Notably, this finding reflects the high number of X-ray structures that show a fully open A-loop conformation only in the presence of either DFG-*in* conformations or the SH2 domain bound to KD (see Table S1).

The HRD motif as the central hub of the signaling network in Abl. Importantly, we observed a relationship between the spatial orientation of the HRD motif in Abl and the specific configuration adopted by the A-loop. In the SH2-KD_{out} and KD_{in} simulations, the HRD motif conserved the *native* conformation, in which the R381 side chain pointed toward Y412, stabilizing the fully open A-loop conformation.¹³³ Conversely, in KD_{out}, the H380 rotated its side chain (Figure S5), causing a local shift of the HRD motif and the partial closure of the A-loop, during which the Y412 gradually adopted *Conf-2* and *Conf-3*. Consequently, R381 ϕ also rotated, causing the HRD backbone flip and the *Conf-3*→*Conf-4* transition (Figure 3.4 and Figure 3.5). In the absence of this R381 ϕ rotation, as observed in additional KD_{out} MD replicas, the HRD motif returned to the *native* conformation through a spontaneous *Conf-3*→*Conf-1* transition (Figure S7).

These results indicate that the spatial orientation of the HRD motif seems therefore critical to understanding A-loop plasticity in Abl. The HRD motif directly connects with the rigid α F helix via two H-bonds that constitute the well-conserved EPK-ELK H-bond network^{139, 147, 148} (Figure 3.1). While the centrality of the HRD motif suggests its role in preserving the Abl folding,¹²⁶ our simulations also suggest the HRD motif's crucial role in modulating a complex network of interactions that, when broken, lead to major rearrangements of the A-loop. In fact, in both the SH2-KD_{out} and KD_{in} systems, the HRD *native* configuration contributed to the preservation of the EPK-ELK H-bond network throughout the trajectories. Conversely, in KD_{out}, the HRD peptide flip and the simultaneous rupture of the EPK-ELK H-bond network trigger the consequent A-loop rearrangement in *Conf-4*.

As expected, in the additional KD_{out} replicas where the HRD motif returned to its *native* conformation, the EPK-ELK H-bond network was also re-formed, as in *Conf-1* state. These results indicate that the *native* HRD motif conformation helps to maintain the proper assembly of the EPK-ELK H-bond network as previously suggested by Oruganty *et al.*¹²⁶ This in turn seems to affect the A-loop conformation, which is critical for Abl function. This was also confirmed by our simulations of the point mutation D440A (i.e. the D440A-KD_{out} system), in which the loss of the EPK-ELK H-bond network and concomitant destabilization of the HRD motif lead to consequent significant reorganization of the A-loop conformational state (see Figure S8 and Figure S9). Interestingly, this finding is in agreement with experimental studies performed on Aurora Kinase, where the point mutation of the well-conserved D311 of the α F helix has been observed to prevent the trans-autophosphorylation of the A-loop threonine (T288).¹²⁶

In addition to the EPK-ELK H-bond network, we observed one key H-bond connecting the DFG and HRD motifs (i.e. Hb[H₃₈₀-D₄₀₀], Figure 3.1), which further helped stabilize the HRD in the *native* conformation and, in turn, the fully open conformation of the A-loop in Abl.^{139, 147} In KD_{in}, Hb[H₃₈₀-D₄₀₀] was formed and preserved well, locking the H380 side chain in its crystallographic orientation. In contrast, in KD_{out}, the Hb[H₃₈₀-D₄₀₀] was not formed because of the DFG-*out* initial conformation, but the H380 side chain interacted with the DFG-1 residue A399. This H-bond (i.e. Hb[H₃₈₀-A₃₉₉]), however, did not preserve the *native* HRD conformation. In fact, Hb[H₃₈₀-A₃₉₉] increased from ~ 3 Å to ~ 7 Å during MD simulations, causing

the rotation of the H380 side chain and the consequent destabilization of the HRD motif, which explains the partial closure of the A-loop. Remarkably, the presence of the SH2 domain in SH2-KD_{out} prevented this destabilization, despite the absence of Hb[H₃₈₀-D₄₀₀]. In SH2-KD_{out}, in fact, Hb[H₃₈₀-A₃₉₉] remained stable, with the HRD motif steady in its *native* conformation.

Taken together, these findings suggest that the HRD motif stabilization is highly interconnected with the preservation of both its H-bond with the DFG motif and the EPK-ELK H-bond network. In turn, this structured H-bond network is critical in modulating the A-loop spatial conformations, and favor its fully open state. The active DFG-*in* conformation of Abl promotes the correct assembly of this H-bond network, whereas the inactive DFG-*out* conformation impairs the proper DFG/HRD interaction, resulting in an overall A-loop destabilization. Finally, the SH2 domain offsets the destabilizing effects provoked by the DFG-*out* motif, stabilizing the fully open conformation of the A-loop in which the Y412 is prone to be phosphorylated (i.e. *Conf-I*, see Figure 3.7).

Finally, it is worth noting that the orientation of the HRD motif affects also the R-spine assembly, which is known to be important for the regulation of kinase activity.^{57, 58} For example, in KD_{out}, the H380 side chain rotation and the HRD peptide flip disrupt the starting configuration of the R-spine, resulting in the A-loop conformational rearrangement. Conversely, in SH2-KD_{out} and in KD_{in} systems, where the HRD motif did not undergo conformational rearrangements, the crystallographic alignment of the R-spine was preserved, with the A-loop that stably maintained in the fully open conformation. Taken together, these results suggest that the R-spine might be involved in regulating the A-loop orientation, being integral part of the signal that goes from the SH2 domain to the A-loop, passing from the β 3- α C loop, a KD portion found to be important to connect the SH2 and KD.¹²⁹ Clearly, this point merits further investigations in order to clarify better the possible connection between the R-spine and signal propagation for modulating the A-loop spatial localization.

Extension to other protein kinases. Structural data of Aurora, Src, and Hck kinases also support the HRD motif as being a key player in modulating the A-loop folding state. As with Abl, the trans-autophosphorylation of T288 in Aurora and Y416 in Src and Hck is necessary to stabilize the A-loop in the fully open

conformation, generating optimal catalytic activity.^{149, 150, 151, 152} In this conformational state, the A-loop is properly stabilized by the electrostatic interaction between the phosphorylated residue and the arginine of the HRD motif, which adopts the same *native* conformation observed in Abl. In detail, the HRD arginine points toward the phosphorylatable residue (T288/Y416) and the EPK-ELK H-bond network and the DFG/HRD interaction are properly formed (see PDBid 1OL5, 3HA6, 3E5A, 3DQW, and 2HK5).^{153, 154, 155, 156} However, different X-ray structures of inactive states of Aurora, Src, and Hck captured different A-loop intermediates, with disordered conformations where the phosphorylatable T288/Y416 was far from the arginine of the HRD motif (PDBid 3DJ5, 3DJ6, 2J50, 1FMK, 2HCK and 1AD5).^{157, 158, 159} Remarkably, in all these structures, the HRD motif was flipped, indicating that preserving the *native* HRD conformation might also be crucial to maintaining the fully open A-loop state in other protein kinases.

Overall, these data indicate that, in Abl, the highly structured H-bond network centered on the HRD motif likely represents an allosteric switch for modulating A-loop spatial orientation. In fact, in our simulations, the rupture of this well-arranged H-bond network is clearly coupled with the destabilization of the fully open A-loop conformation. This induces major A-loop rearrangements, with formation of an ensemble of A-loop conformations that are poorly prone to undergoing phosphorylation. Structural data analysis of other tyrosine and serine/threonine kinases further supports the idea that proper HRD motif orientation and preservation of the structured H-bond network centered on it are critical to maintaining the A-loop in the fully open conformation. Thus, these results suggest that specific key structural components of KD in Abl (namely the DFG motif, the HRD motif, and the EPK-ELK H-bond network) act cooperatively to regulate the A-loop dynamics and spatial arrangement. In our simulations, the central hub of this extended signal network was the HRD motif, which may thus be the core of the signal transduction mechanism for (allosteric) modulation in Abl. Based on the structural similarity of Abl with several other kinases, it seems reasonable to extend this hypothesis to other protein kinases such as Aurora, Src, and Hck kinases, where the conserved HRD motif could act similarly to modulate the signal transduction pathways within the KD domain.^{57, 148} The SH2 domain is an allosteric control mechanism for this complex signaling network of A-loop spatial orientation and Abl activity.

Thus, these results augment the interpretation of the existing Abl experimental data and may help the rational design of new Abl binders that interfere directly or allosterically with this signaling network.^{160, 161, 162}

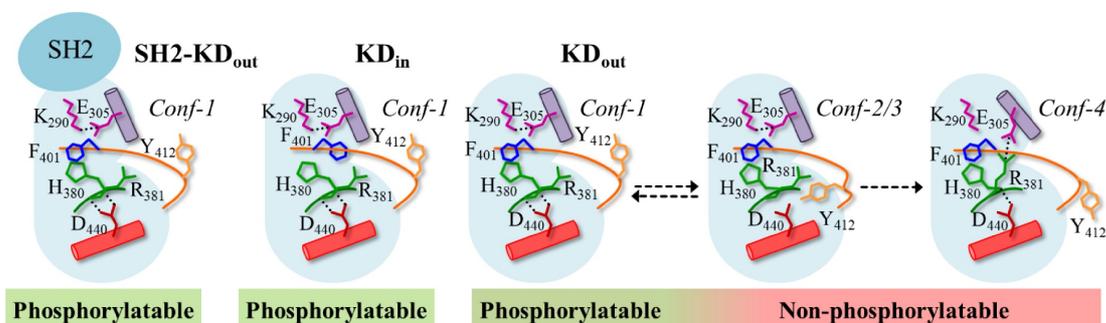


Figure 3.7. Schematic representation of the Y412 conformations observed during our MD simulations of the SH2-KD_{out} (left), KD_{in} (middle), and KD_{out} (right) systems. All simulations started in Conf-1 state, where the A-loop was in the fully open conformation and prone to be phosphorylated. In both SH2-KD_{out} and KD_{in}, the A-loop remained in the initial phosphorylatable Conf-1 state. Conversely, in KD_{out}, the A-loop conformation was partially closed during the MD, visiting Conf-2 and Conf-3. Multiple MD replicas from these metastable conformational states showed that the A-loop can re-open either toward a nonphosphorylatable intermediate Conf-4 state or toward the starting phosphorylatable, fully open Conf-1 state.

METHODS

Structural model and classical MD simulations. We used 5 model systems, namely: 1) the Abl kinase domain (KD) in the DFG-*out* conformation and in complex with the SH2 modulatory subunit (SH2-KD_{out}); 2) the DFG-*out* KD, without the SH2 domain (KD_{out}); 3) the DFG-*in* KD (KD_{in}); 4) the DFG-*out* KD with the D440A mutation (D440A-KD_{out}); and 5) the DFG-*out* KD with the A-loop in the closed conformation (closed-KD_{out}).

The SH2-KD_{out} system was constructed using the X-ray structure of Abl in the DFG-*out* conformation bound to the SH2 modulatory domain in the ‘top-hat’ conformation (PDBid 1OPL chain B).⁷² Both the SH2-KD linker and the terminal portion of the α I helix, which are missing in the X-ray, were modeled using Prime software of Maestro suite (Figure S11). The resulting model is consistent with the Abl sequence length used in the experiments of Lamontanara *et al.* (see SI for details).¹²⁸ The KD_{out} system was built using the same structure used for the SH2-

KD_{out} model, after removing the SH2 modulatory domain. The KD_{in} system was constructed using the X-ray structure of Abl in the DFG-*in* conformation (PDBid 2F4J).¹⁶³ The D440A-KD_{out} system was constructed using the KD_{out} model, introducing the D440 mutation. The closed-KD_{out} system was constructed using the X-ray structure of Abl kinase bound to imatinib and myristic acid (PDBid 1OPJ chain B). To ensure that the removal of myristic acid does not significantly perturb the structure of the myristate pocket, we monitored the behavior of this hydrophobic cavity along the MD simulations (see SI for further details). For all systems, the co-crystallized ligands were removed. Each system was solvated with TIP3P water molecules¹⁶⁴ and neutralized with Na⁺ ions. The total number of atoms was ~76000 (SH2-KD_{out}), ~45-50000 for each of the other systems (see SI for more details). The proteins were parameterized using the Amber99SB force field¹⁶⁵ (see SI for details on the MD set up). Overall, we collected MD simulations of ~1.8 μ s for SH2-KD_{out}, ~2.6 μ s for KD_{out}, 1.1 μ s for both Replica 1-KD_{out} and 2-KD_{out}, and ~1.3 μ s for KD_{in}, D440A-KD_{out} and closed-KD_{out}, for a total of over ~10.5 μ s of MD.

Analysis of structural data from MD simulations. The root-mean-square deviation (RMSD) of the protein backbone after the equilibration time (~100 ns) with respect to the crystal structures was used to evaluate the stability of the systems during the MD simulations and perform comparative analyses. The principal component analysis (PCA) was performed to reduce the dimensional space of the data sets from MD trajectories, highlighting the principal motions of the systems.¹⁶⁶

Experimental evidence suggests that Y412, situated along the A-loop, is efficiently trans-autophosphorylated when the A-loop assumes the fully open conformation (as in the X-ray structure PDBid 1OPL chain B), whereas it is less prone to be phosphorylated in closed conformations (as in the PDBid 1OPJ). In our MD simulations, we adopted two parameters to define the conformational rearrangement and spatial delocalization of Y412 along the A-loop: 1) the RMSD of Y412 with respect to the reference X-ray structure 1OPL chain B (hereafter RMSD_{Y412}); 2) the distance between the phenolic oxygen of Y412 and the center of mass of the carboxylic oxygens of D440 located in the rigid α F helix ($d[Y_{412}-D_{440}]$, see Figure 3.4).

Based on this metric, we obtained four clusters that correspond to a given Y412 conformation (Figure 3.4): i) Cluster 1 corresponds to the ensemble of Y412

conformations where the A-loop is in the fully open conformation, as in the 1OPL X-ray structure, and the Y412 is prone to be trans-autophosphorylated (*Conf-1*). The minimum point of this cluster is located at $\text{RMSD}_{\text{Y412}} = \sim 1.5 \text{ \AA}$ and $d[\text{Y}_{412}\text{-D}_{440}] = \sim 13 \text{ \AA}$. ii) Cluster 2 corresponds to those conformations in which the A-loop is partially closed and, consequently, the Y412 is not prone to be trans-autophosphorylated (*Conf-2*). This cluster is centered at $\text{RMSD}_{\text{Y412}} = \sim 3.5 \text{ \AA}$ and $d[\text{Y}_{412}\text{-D}_{440}] = \sim 5.5 \text{ \AA}$. iii) Cluster 3, as in *Conf-2*, represents the conformations in which the A-loop is partially closed, but this time the Y412 directly interacts with the D440 via an H-bond (*Conf-3*). Here too, the Y412 assumes a non-optimal position for trans-autophosphorylation. The minimum point of this cluster is located at $\text{RMSD}_{\text{Y412}} = \sim 4.5 \text{ \AA}$ and $d[\text{Y}_{412}\text{-D}_{440}] = \sim 3 \text{ \AA}$. iiiii) Cluster 4 corresponds to those conformations in which the A-loop is open and the Y412 is flipped with respect to the conformation seen in *Conf-1*. Thus, Y412 assumes a non-optimal conformation for trans-autophosphorylation (*Conf-4*). This cluster is centered at $\text{RMSD}_{\text{Y412}} = \sim 11 \text{ \AA}$ and $d[\text{Y}_{412}\text{-D}_{440}] = \sim 15.5 \text{ \AA}$. These cutoffs were chosen to optimally separate the frequency distributions of the two variables (see SI).

A recent study has demonstrated that the EPK-ELK H-bond network formation could be related to a proper catalytic activity of kinases.^{126, 167} In Abl, the EPK-ELK H-bond network comprises two H-bonds between the oxygen atoms of the D440 carboxylate group in the αF helix and the backbone amide of both the H380 and R381 residues of the HRD motif (hereafter referred to as $\text{Hb}[\text{H}_{380}\text{-D}_{440}]$ and $\text{Hb}[\text{R}_{381}\text{-D}_{440}]$, respectively). In Abl, the proper assembly of the EPK-ELK H-bond network depends on the backbone torsional angle ϕ of R381 ($\text{R381}\phi$), which favors the interaction between the R381 side chain and the phosphorylated Y412, further locking the open A-loop conformation.¹²¹ Thus, we consider Abl to be fully active only when the A-loop is fully open (i.e. prone to be trans-autophosphorylated), the DFG motif assumes the *in* conformation, and the EPK-ELK H-bond network is formed.

Multiple walkers metadynamics. Here, we are interested in computing the free energy surface (FES) of the A-loop conformational transition from the fully open conformations, in which the Y412 is prone to be trans-autophosphorylated as in the crystal 1OPL chain B, to the partially closed conformations, in which the Y412 is less prone to be phosphorylated, as observed in our unbiased MD simulations. These

two states are described by the distance $d[Y_{412}-D_{440}]$, i.e. the employed collective variable (CV). However, MD simulations show that the A-loop undergoes partial unfolding at large CV values, adopting unstructured states as those observed in KD_{out} and in 1OPJ simulations (see KD_{out} paragraph and Figure S12). In our study, these unstructured states of the A-loop are of no interest and, thus, should be avoided in sampling the conformational space.

To do so, we considered our system having a bistable energy profile as that reported in Figure 3.8, where the basin A gathers the structured A-loop conformations (both prone and non-prone to be trans-autophosphorylated), while the basin B includes the unstructured A-loop conformations, i.e., structures of no interest. Our goal is therefore to sample as much as possible the A-loop conformational states in basin A, possibly avoiding the sampling of basin B. To this aim, one may simply add a restraining potential at the CV value that separates A from B. However, this simple approach is often related to issues that can be far from trivial. For example, a soft restraining potential would not prevent the exploration of basin B. On the other hand, increasing the restraining potential may induce significant structural distortions, especially in systems characterized by a large flexibility, as in our case. Moreover, the use of walls may originate problems of hysteresis.

To overcome these potential issues, here we developed and applied a simple new approach that slightly differs from the use of potential walls. Our approach is based on the confinement of the multiple walkers metadynamics runs. That is, when a metadynamic run (a “walker”) hits a specified CV boundary value (indicating that the system is entering in the basin B), the simulation is restarted from conformations stored in a pool P of frames saved during the previous metadynamics simulations in basin A. In this way, metadynamics continue the exploration of the relevant basin by improving the sampling of transversal and often rather slow degrees of freedom. In our simple bistable system A and B, this approach proceeds as follows: 1) the simulation is monitored every t ps and the state of the system (in either A or B) is determined; 2) if the system is in A, the simulation continues and the current conformation is stored in the pool P of restart structures; 3) if the system happens to be close to fall into B, it is placed back into the basin A, by randomly choosing a restart conformation from P . The whole procedure represents a specific realization of

the technique of multiple walkers introduced by Raiteri *et al.*¹⁶⁸ The walkers are asynchronous, as a new walker only starts when the preceding crossed the boundary value. The use of multiple walkers can be very powerful to improve the convergence of the free energy in CV subspaces of complex biomolecular systems, as also proposed recently for protein-ligand complexes.

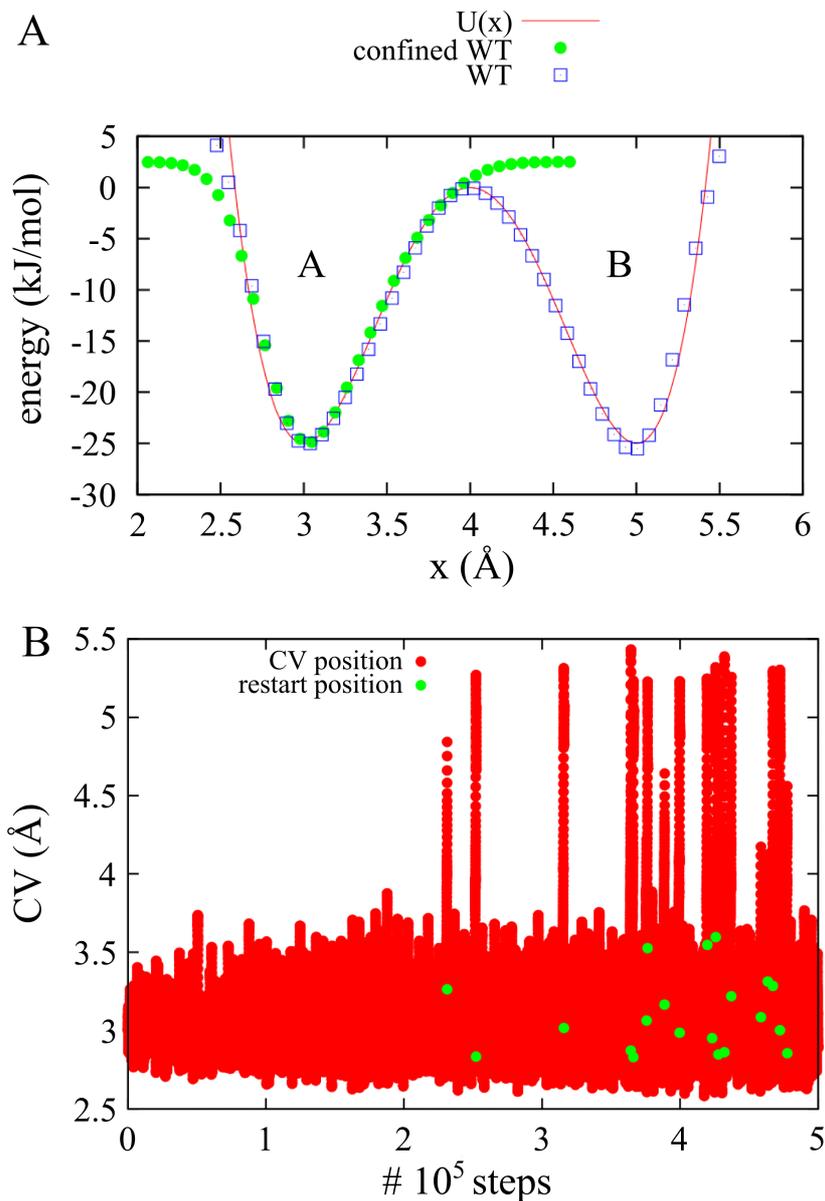


Figure 3.8. A) 1D bistable potential with form $U(x) = a[(1/2)(x - 4)^2 - (1/4)(x - 4)^4]$ with $a=100$ kJ/mol. The basin A is defined by $CV < 4\text{Å}$ while basin B is defined as $CV \geq 4\text{Å}$. The free energy obtained with well-tempered (WT) metadynamics is reported together with that obtained by means of the confinement procedure. B) Behavior of the CV in the confined simulation. Green dots indicate the position of the CV at the first step of the restarted system.

Importantly, we validated our *confined multiple walkers metadynamics* on a simple 1D bistable spring with potential energy $U(x)$ as shown in Figure 3.8. We performed well-tempered metadynamics on this system by biasing the spring length, using a hill height of 0.2 kJ/mol, a hill width of 0.2 Å, a fictitious CV temperature of 1500 K, a system temperature of 300 K, a deposition rate of 1 ps⁻¹ and the total simulation length 2 million steps. This allows reproducing the free energy profile in both basins, A and B. The restarting procedure is applied to restrict the sampling within basin A. The boundary CV value is set to 4 Å. As it can be noticed in Figure 3.8, the free energy profile obtained for basin A converges to the profile of the standard well-tempered simulation after few restarts, and a total simulation length of 500,000 steps. Notably, restart conformations can be extracted with equal probabilities, as long as the walker spends enough time in the basin of interest, losing memory of its initial position (for a discussion on the related timescales see ref. ¹⁶⁸).

In the case of Abl kinase, we adopted well-tempered metadynamics (WT-MetaD)^{106, 146} and two walkers for each system (i.e. KD_{out}, KD_{out}, and KD_{in}) that started from two different A-loop conformations: i) the fully open *Conf-1* (d[Y₄₁₂-D₄₄₀] ~12 Å); ii) the partially closed *Conf-3* (d[Y₄₁₂-D₄₄₀] ~3 Å) – see SI for details. The confined simulations are monitored every 1 ns. The FES is determined as a function of d[Y₄₁₂-D₄₄₀], which is the CV that identifies the partial closure of the A-loop. Finally, it is worth to mention that we employed a CV value of 17 Å for our confined multiple walkers metadynamics (see Figure S12). This allowed us to avoid the well-known problem related to possible distortion of the free energy profile around the boundary, in a CV interval of approximately the same size of the gaussian width used.^{169, 170, 171, 172}

Supporting Information

Details on systems preparation, protocols employed for MD simulations, details of analysis of data, and tables and figures referenced in the text are available at Appendix A.

Chapter 4

Automated analysis of binding pocket
dynamics and allosteric networks in
proteins

ABSTRACT

Detection and characterization of binding pockets in proteins is crucial to study biological regulation, and perform drug design. This is because endogenous ligands and drugs modulate the target's function through their binding to specific functional pockets, sometimes also connected through allosteric networks. In these regards, ever-increasing molecular dynamics (MD) simulations are nowadays used routinely to investigate the spatiotemporal evolution of biomolecules, and characterize their innate dynamics and plastic behavior within the cell. Yet, there is no computational tool for the automatic analysis of binding pocket dynamics and allosteric signaling networks in proteins, as represented by an MD-generated equilibrium ensemble of structures. Here, we address this need presenting a fully automated method that implements an original algorithm to analyze extended MD trajectories, without any *a priori* information on the existence of pockets and allosteric networks. We exemplify the method's capability to efficiently analyze microsecond-long MD simulations, and match experimental data, of a few selected pharmaceutically relevant proteins (PNP, A_{2A} and Abl kinase). These analyses exemplify the potential of this method, including its prospective use for binding pocket characterization in structure-based drug design.

INTRODUCTION

Binding pockets are often crucial for modulating the function of biomolecules, such as those in protein enzymes and ion channels. For this reason, small molecule drugs exert their action through their binding to a functional pocket of the protein target(s). Binding to functional pockets is, thus, at the very basis of the beneficial pharmacological effect of drugs. Detection and characterization of these functional binding pockets is therefore of paramount importance for biochemistry and drug discovery.

In this respect, molecular dynamics (MD) is nowadays routinely used to characterize the spatiotemporal evolution of biomolecules. These MD simulations now span from hundreds of nanoseconds to even few milliseconds, therefore allowing to explore exhaustively the dynamics and plasticity of large protein systems.²⁹ This means, importantly, that MD can also capture the appearance, evolution and structural modifications of binding pockets. Notably, these include also those superficial and shallow transient cavities that are often involved in protein function through their interaction with a small substrate or another partner protein. However, in MD, these events are described by large structural ensembles forming increasingly extended trajectories, which nowadays can reach hundreds of Gigabytes and, soon, tens of Terabytes and more. Therefore, a major need today is to develop algorithms capable to extract automatically the embedded information from these big MD-generated data sets and generate intelligible reports on the spatiotemporal evolution of the targeted protein, including its potentially druggable binding pockets.

Already, there are a number of different algorithms for the detection of protein binding pockets in static structures. Some of these rely on the Voronoi diagrams (MolAxis,¹⁷³ MOLE¹⁷⁴), grids (POCKET,¹⁷⁵ PocketFinder¹⁷⁶), or the molecular surfaces and probes (HOLE,¹⁷⁷ SURFNET¹⁷⁸). Later, other advanced algorithms were derived from these, including NanoShaper,¹⁷⁹ which is our algorithm based on the rigorous Alpha Shapes theory.^{180, 6, 48, 4} All these algorithms are instrumental to the analysis of an MD-generated conformational ensemble of structures and their binding pockets, usually performed via structural alignment, such as in MDpocket,⁴⁹ PocketAnalyzer^{PCA},¹⁸¹ Epock,¹⁸² Trj_cavity¹⁸³ and TRAPP.¹⁸⁴ However, the resulting information may be affected by the specific reference structure used for the alignment. In alternative, atom-based algorithms like PROVAR¹⁸⁵ and EPOS^{BP},¹⁸⁶

avoid the alignment step. Nevertheless, most of these methods limit their analysis to the *a priori*-defined pocket of interest, while none of them is able to recognize possible signal transmission networks formed by distal allosteric pockets. Automatic detection of allosteric signals in proteins remains challenging, with methods based on normal mode,¹⁸⁷ community network analysis,^{46, 130} coevolution analysis²³ and machine learning techniques.¹⁸⁸ These methods, however, do not allow an automatic analysis of the spatiotemporal evolution of pockets, and their communication network, as investigated by extended MD simulations.

In this context, here we present an original algorithm to analyze extended MD trajectories, which automatically execute **a)** an alignment-independent identification of all the pockets on the protein surface, **b)** a quantification and visualization of the volume and surface area of all the pockets found in the protein, for each structural frame forming the MD trajectory, **c)** a history graph of interactions between pockets, described as merging and splitting events, **d)** a detection of allosteric signal transmission networks along the protein surface, defined as interconnected pocket motions (see theory section).

We demonstrate the power of this new algorithm through the unprecedented analysis of the structure-dynamic-function relationship automatically extracted from extended MD simulations of a few selected test cases of pharmaceutical relevance, namely: i) the purine nucleoside phosphorylase (PNP) enzyme, ii) the adenosinic receptor (A_{2A}) simulated in POPC membrane and ii) the Abelson (Abl) kinase. Taken together, these examples illustrate the potential large implications of such powerful analysis, including its use for structure-based drug design.

RESULTS

Protein binding pocket detection and dynamics. First, we tested the ability of our new algorithm implemented in Pocketron to identify and monitor protein pocket dynamics. Toward this end, we first used a ~700 nanosecond long MD simulations of the purine nucleoside phosphorylase (PNP), a homotrimeric enzyme that is involved in the purine metabolism and T-cell function.^{189, 190} PNP inhibition is a strategy to treat T-cell mediated diseases such as leukemia and lymphoma.¹⁹¹

Pocketron identified a total of 22 pockets that exist for more than 30% of the overall simulations describing the dynamic docking of DADMe-ImmH into PNP.

Among the most persistent ones, Pocketron identified three large pockets, namely pocketID (pID) 4, pID 9 and pID 12 (Figure SI 1). These have a persistency of ~79% for pID 4 and ~96% for both pID 9 and pID 12. These three pockets are at the edge between the adjacent PNP monomers (see Figure 4.1). Each of them corresponds to the known orthosteric binding site that is targeted by PNP inhibitors, such as Immucillin-H and DADMe-ImmH.^{192, 193} These pockets have an average volume of 367.0, 893.9 and 976.6 \AA^3 , respectively. These volumes are larger than in the holo PNP crystal, where these pockets are 332.1, 326.7 and 408.6 \AA^3 , respectively. This indicates that the orthosteric pockets, possess an elevated plasticity. Notably, the reduced volume of pID 4 is explained by the existence of a nearby stable pocket, pID 13, which has a persistency of 94% and an average volume of 313.5 \AA^3 . These two pockets merge for ~26% of the simulations time, resulting in a single larger orthosteric binding pocket having an average volume of 676.9 \AA^3 , which is comparable with the other two orthosteric pockets pID 9 and pID 12. Furthermore, our algorithm detected two smaller pockets, pID 3 and pID 14, located in proximity of the orthosteric binding pockets pID 4 and pID 12 (Figure 4.1 and Figure SI 1). These pockets are present for the 71% and 73% of the simulation time and have an average value of 132.7 \AA^3 and 144.4 \AA^3 , respectively, in agreement with their value in the PNP X-ray structure (202.9 and 232.3 \AA^3 , respectively). Interestingly, these pockets have been shown to be a pre-binding site in which the inhibitor DADMe-ImmH transiently binds before accessing on of the orthosteric sites. Notably, this pre-binding site was not found in proximity of the orthosteric pocket pID 9.¹⁹⁴ Likely, this is because during our simulations, pID 9 already contains the pre-binding site, as suggested by its larger volume compared to pID 4 and pID 12.

Another pocket, namely pID 22, is detected at the center of the trimerization interface (Figure SI 1), with an average volume of 397.9 \AA^3 (248.1 \AA^3 in the X-ray structure) and a time persistency of 88%. Intriguingly, this pocket shows a marked communication pattern with the orthosteric binding site of each monomer during the MD simulations (Figure SI 2). This suggests a signaling transmission network among the different monomers mediated by this common interface pocket, possibly explaining the “one-third-the-site” observation of Schramm and coworkers¹⁹⁵ (see Discussion).

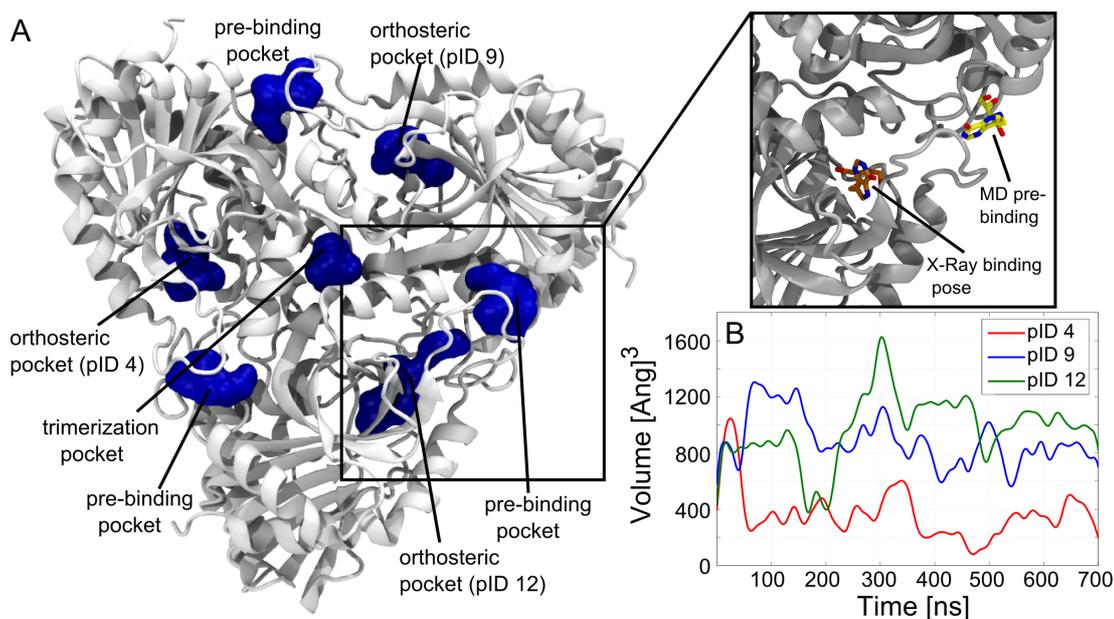


Figure 4.1. A) Localization of the main pockets computed for the PNP X-ray structure 3K8O. On the right, the orthosteric ligand DADMe-ImmH is positioned in the orthosteric pocket (orange molecule) as in the X-ray structure and in the pre-binding pocket (yellow molecule) as in our MD simulations. B) Volume along time of the three orthosteric pockets pID 4, pID 9 and pID 12.

Identification of interaction patterns among pockets in proteins. Our algorithm detects pocket interactions as pocket splitting and merging along an MD simulation (see Theory). Here, we demonstrate such feature on 1) the Adenosinic Receptor A_{2A} , a G-protein coupled receptor (GPCR) that regulate myocardial blood flow, the release of glutamate and dopamine, and that is a promising target for pain, depression and neurological diseases as Parkinson; 2) the Abelson (Abl) tyrosine kinase, a protein involved in cell growth and survival. Abl kinase is a validated target for the treatment of several types of cancer.⁶⁶ Importantly, both proteins present multiple binding pockets that are largely characterized by biochemical and structural investigations.

Adenosinic Receptor A_{2A} . The analysis performed on 100-nanosecond long A_{2A} trajectory revealed the presence of 13 pockets, which are present for more than 30% of the simulations. Among them, pID 15 and pID 22 coincide with the already known and structurally characterized orthosteric and allosteric binding sites, respectively. In particular, pID 15 is located in the extracellular side of the receptor

and is targeted by both agonist and antagonist drugs, as adenosine¹⁹⁶ and ZM241385.¹⁹⁷ Instead, pID 22 is located in the core of the transmembrane (TM) bundle and is normally occupied by a sodium ion that can be displaced by allosteric drugs as amiloride and HMA (see Figure 4.2).^{198, 26} Our analysis revealed that pID 15 and pID 22 communicate during the MD simulations through merging and splitting events. This means that these pockets share a set of residues at the interface between them. These residues are Val 84, Ala 88, Phe 242 and Trp 246 (see Figure 4.2 A), which allow a cross-talk between these pockets. Interestingly, MD simulations and experimental studies have demonstrated ligand-dependent A_{2A} signaling involving an allosteric effect through these two pockets.^{199, 198, 200}

Although the two pockets are physically well separated during most of the simulations, Pocketron identified and visualized when the two pockets underwent either merge or split events. For example, we show frame 844 – out of a thousand frames analyzed – where the larger orthosteric pocket pID 15 incorporates the allosteric pID 22 (see Figure 4.2 B). Remarkably, when the two pockets are merged together, they allow the formation of a small channel that can permit the passage of the sodium from the extracellular side to the allosteric pocket, as suggested by thermal stability studies¹⁹⁸ and MD simulations.²⁰¹ In this context, the shared residues at the edge between the two pocket (again, Val 84, Ala 88, Phe 242 and Trp 246) might act as a gate of the channel, mediating its opening, closure, and thus modulating sodium entrance. In fact, it has been well established that the conformation flexibility of Trp 246, known also as “toggle switch” residue, is associated with the activation mechanism of GPCRs.^{202, 203, 204, 205} In summary, these results show that Pocketron describes interaction patterns between orthosteric and allosteric pockets, in agreement with experimental data.

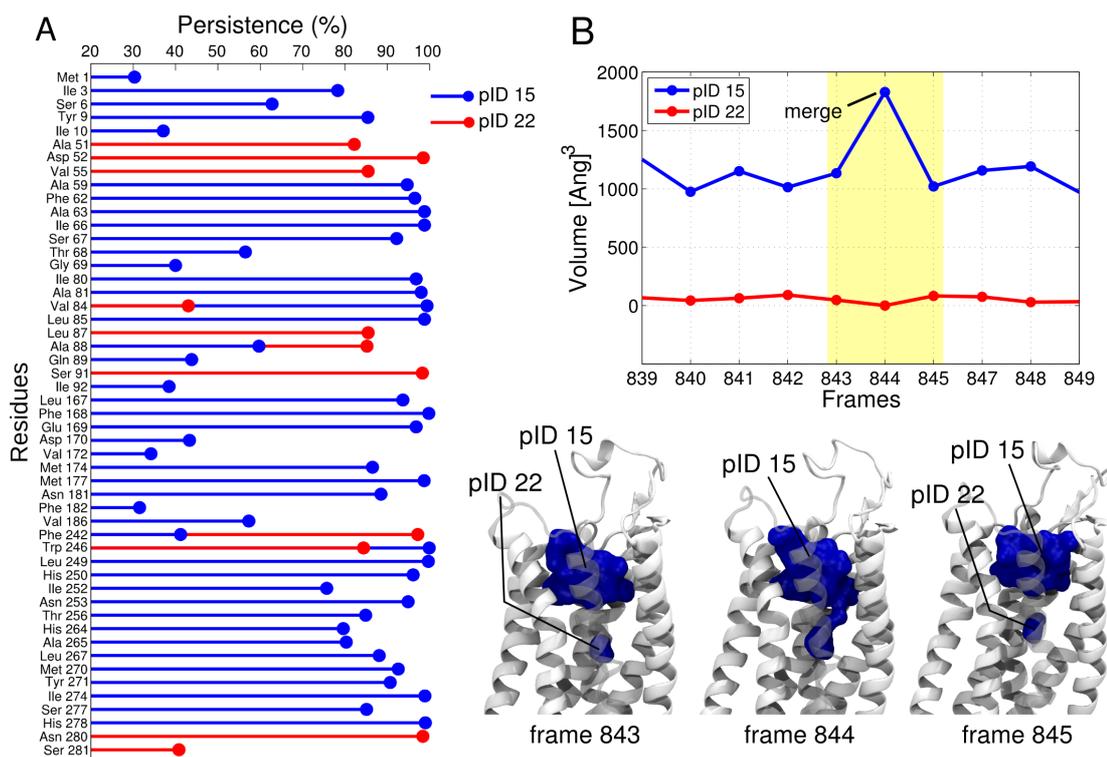


Figure 4.2. A) Time persistency of residues that define the orthosteric pocket (blue stem, pID 15) and the allosteric pocket (red stem, pID 22). B) Volume plot of pID 15 and pID 22 in selected frames and representation of merging and splitting events.

Abelson (Abl) tyrosine kinase. Here, we analyzed the MD simulations of the kinase domain (KD) of Abl, both in the catalytically active DFG-in (KD_{in}) and inactive DFG-out (KD_{out}) conformations (see Methods for details). Among the most persistent pockets in KD_{in} , Pocketron successfully identified the large ATP pocket (pID 5). Besides pID 5, Pocketron also found another nearby pocket in between the DFG motif and the αC helix (pID 3), which is known to be an allosteric site in KD. In fact, the two pockets, pID 5 and pID 3, do actually communicate in our MD simulations, having a large set of shared residues listed in Table SI 1. For $\sim 63\%$ of the simulations, they remain two separated pockets with volume 438.7 \AA^3 and 166.7 \AA^3 . However, for the $\sim 35\%$ of the simulations, the two pockets merge into a single large pocket having an average volume of 497.4 \AA^3 (see Figure 4.3).

Also in KD_{out} , Pocketron identified the ATP pocket pID 5 and the nearby allosteric pocket pID 3 having an average volume of 338.6 \AA^3 and 221.8 \AA^3 , respectively. However, this time an additional pocket was identified, pID 28, which is located in between the P-loop and the $\beta 3$ - αC -helix loop, and which has a smaller volume of 150.5 \AA^3 . Notably, the region containing this new pocket has been

targeted by a new class of diphenylamine-derived allosteric inhibitors in MEK 1 and MEK 2 kinases.²⁰⁶ Also here, the three pockets pID 3, pID 5 and pID 28 share a set of residues (listed in table SI 1), which indicates a communication network among them. Interestingly, in this case they exist separately only for ~11% of the overall simulated time, due to more frequent merge events with respect to the catalytically active KD_{in} system (see Figure 4.3).

According to our analysis, in the catalytically active KD_{in} system, the DFG-in conformation limits the cross-talk between the ATP pocket and the nearby pocket pID 3. In contrast, in KD_{out} , the DFG-out conformation favors the communication with the surrounding pockets pID 3 and pID 28. Indeed, the DFG-out conformation has been demonstrated to augment the flexibility of the ATP binding site, which is translated into the loss of KD activity.^{207, 129} Thus, our analysis suggests that the specific DFG conformation, in or out, affect the formation of transient pockets and their communication network, which in turn modulates the overall shape of the ATP binding site and the resulting KD activity.

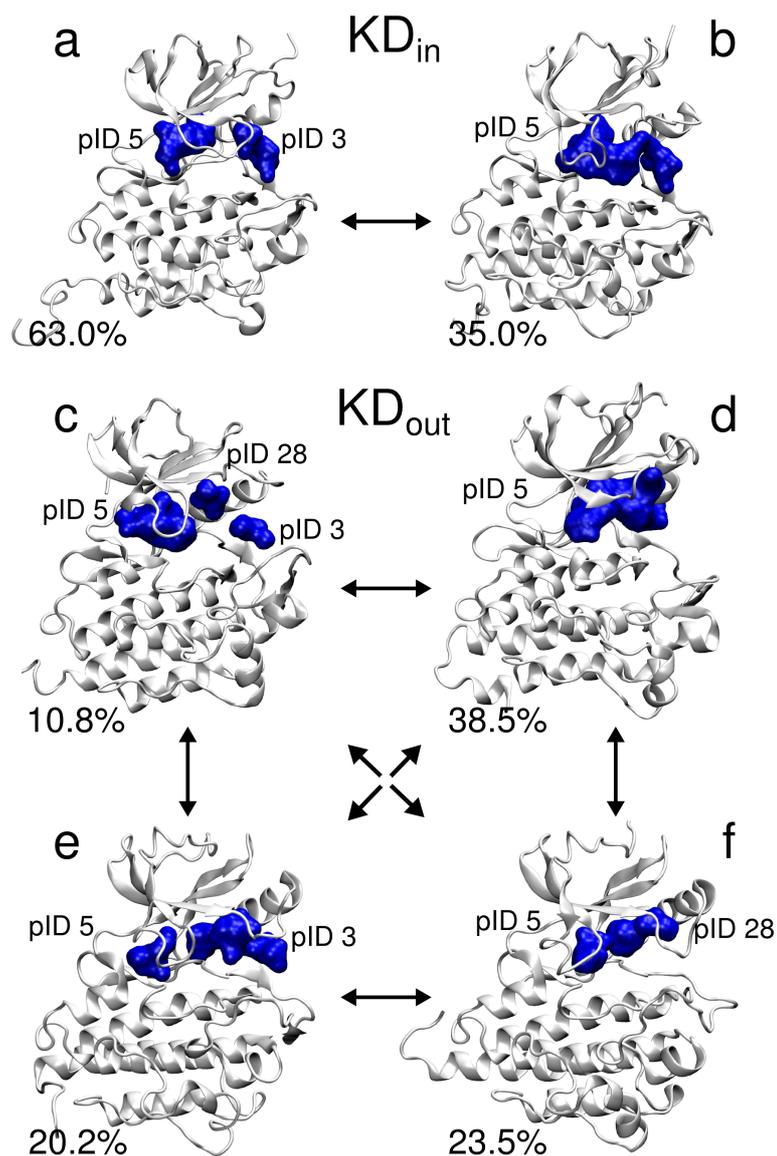


Figure 4.3. Representation of the dynamical behavior of the ATP pockets during the MD simulations of KD_{in} system (a and b) and KD_{out} systems (c-f). In the KD_{in} , in (a) the ATP pocket pID 5 coexists with the near pID 3 for the 63.0% of the simulation, while in (b) pID 5 is the only pocket for the 35.0%. In KD_{out} , in (c) the ATP pocket coexists with pID 3 and pID 28 for the 10.8% of the MD simulation; In (d) the ATP pocket is the only binding site for the 38.5% of the simulations, while in (e) it coexists with pID 3 for the 20.2% and in (f) it coexists with pID 28 for the 23.5% of the simulations.

Allosteric signal transmission networks inferred from pocket dynamics. Here, we tested the capability of our new algorithm to retrieve allosteric signal propagation pathways in proteins as represented by an MD-generated equilibrium ensemble of structures. As a paradigmatic example, we investigated the Abl kinase, which is one of the most characterized systems in terms of molecular determinants for protein

allostery.^{129, 207} We performed a comparative analysis of four different Abl systems: 1) the wild type KD_{in} form, 2) KD bound to myristate (Myr/KD_{in}), 3) the KD apo form of the T315I mutation (T315I-KD_{in}), 4) system T315I-KD_{in} bound to myristate (Myr/T315I-KD_{in}).

In all systems, our analysis detected both the orthosteric ATP pocket and the allosteric myristate pocket, which is located in the C-lobe of the KD.¹¹⁴ Also, the analysis of the MD frames returned an intensive interconnection of atoms across pockets distributed on the protein surface. These interconnections were detected as off-diagonal elements of the square matrix **N** (see Theory). Notably, such a cross-talk establishes a long-range communication network between the distal orthosteric ATP and the allosteric myristate pockets (see Figure 4.4), which is well-known to be central for the allosteric signal propagation in Abl. In particular, in both KD_{in} and Myr/KD_{in}, the ATP and the myristate pockets are linked through a long-range communication network that passes through a number (~3 to 5) of small transient pockets (average volume ~90 Å³, and persistency between 20 and 70%) mainly located in the C-lobe and involving mostly the αG and αI-helix (Figure 4.4). Notably, the location of pockets coincides with that of the new allosteric pockets reported by Shan *et al.* in their study on binding pathways in Src kinase. Notably, in the MD simulations of Abl with the drug resistant T315I point mutation (i.e. T315I-KD_{in} system), the ATP ↔ myristate long-range communication network is actually disrupted (Figure 4.4). The perturbation of such communication pathway detected in T315I-KD_{in} system might explain dysregulation of the T315I kinase form.²⁰⁸ Remarkably, however, in our simulations where the myristate bind to the T315I Abl mutated form (i.e. Myr/T315I-KD_{in} system), the ATP ↔ myristate pockets long-range communication network is completely restored, as in the wild-type system. This result is in line with experimental studies that show that myristate mimetics elicit a structural rearrangement on the ATP pocket, influencing the binding affinity of orthosteric inhibitors.^{114, 209} This MD-based allosteric communication network²¹⁰ could also explain HX MS experiments that further detailed the presence of an allosteric signal propagation pathway from myristate to the ATP pockets, passing through KD structural elements as the αI-helix.²⁰⁹

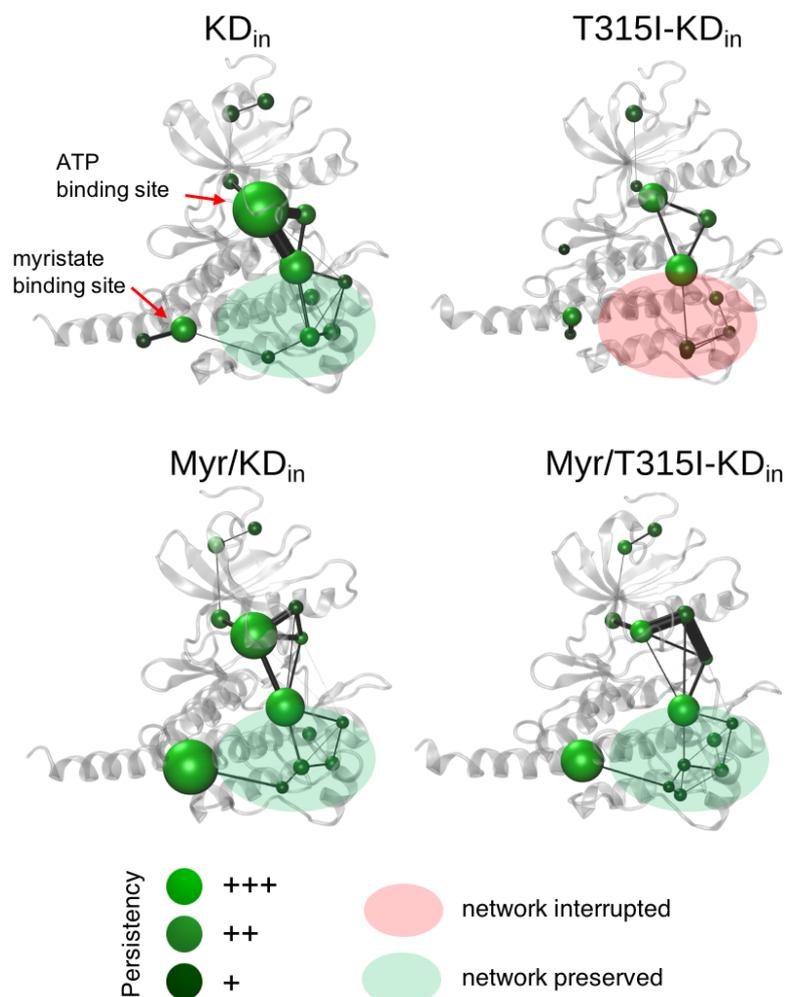


Figure 4.4. Networks of the most persistent pockets found in KD_{in} , $T315I-KD_{in}$, Myr/KD_{in} and $Myr/T315I-KD_{in}$ trajectories. Each pocket (i.e. node of the network) is represented as a green sphere and the different colors indicate the pocket persistency. The pockets are connected through black lines (i.e. edges of the network). The width of each edge is proportional to the communication frequency. The networks connect the ATP binding site and the myristate binding site in all systems except for $T315I-KD_{in}$.

DISCUSSION

In this work, we present a novel algorithm, implemented in a new software named Pocketrone, for the detection of all pockets on a given biomolecular system. This algorithm allows examining extended MD trajectories, monitoring the spatiotemporal evolution of all the cavities, including those transient pockets that form allosteric communication networks to modulate protein function. Notably, this analysis was performed without any *a priori* information on the existence of pockets and communication network. Here, we validate this algorithm using three different

biomolecular systems, namely the Purine Nucleoside Phosphorylase (PNP) enzyme, the Adenosinic Receptor (A_{2A}) and the Abelson (Abl) kinase.

Firstly, Pocketron identified all the experimentally known orthosteric binding sites in the PNP homotrimeric enzyme, namely pID 4, pID 9 and pID 12. Interestingly, these pockets are also among the most persistent and larger ones, indicating that these two parameters might be important to evaluate promising targetable sites and hot spot. In addition, our analysis also identified smaller pockets in proximity of the aforementioned larger binding sites. Notably, these auxiliary pockets were involved in a pre-binding state of DADMe-ImmH, a PNP inhibitor, which was found to transiently interact with this small pocket, before entering into the main orthosteric site.¹⁹⁴ The analysis of the volume variation along time evidenced an elevated plasticity of the orthosteric binding pockets. In fact, the average volume of the three pockets is higher respect to crystallographic value, indicating that during the simulations the pockets undergo to a structural rearrangement which facilitate the binding of DADMe-ImmH inhibitors observed during the simulations. Also, our analysis revealed a marked communication pattern between the orthosteric pocket and nearby pre-binding site pocket which is due to an elevated frequency of merge and split events. Taken together, our findings detailed an elevated plasticity of PNP orthosteric binding sites. Remarkably, this is in line with structural studies that showed a major conformational alteration of two structural motifs of the binding site (i.e. His 257-helix and His 64-loop) upon inhibitor's binding, which affect the size and the shape of binding pocket.²¹¹ Notably, the elevated flexibility of this structural region in PNP has been used recently to demonstrate that MD can be effectively employed to improve virtual screening results. Pocketron also identified a long-range communication pathway among the orthosteric pockets and another pocket detected at the center of the trimerization interface (pID 22). This pathway might indicate the presence of an allosteric signal that connects the far orthosteric pockets, passing via the trimerization interface. This can support the “one-third-the-site” theory of Shramm and collaborators, although further computational and experimental studies are still needed to clarify the allosteric behavior of PNP enzyme. Overall, these results confirm the ability of this new algorithm to detect all pockets on the protein surface, and monitor their conformational flexibility, irrespective of their size, location and extension.

Subsequently, Pocketron was used to monitor local communication network among near binding sites described as pocket merge and split events (see Theory). To do so, we first analyzed the MD trajectory of A_{2A} simulated in POPC membrane. Also here, Pocketron was able to identify the experimentally known pockets, which are the orthosteric binding site, located in proximity of the extracellular side, and the underlying sodium allosteric site. In addition, our analysis showed that the two pockets communicate during simulation, sharing a set of residues located in the interface of the two sites (i.e. Val 84, Ala 88, Phe 242 and Trp 246). Notably, this result is in line with experimental and computational studies that propose an allosteric cross-talk between the two cavities, which is crucial for the modulation of A_{2A} function. In this regard, the plasticity of this region might have a functional role in mediating the allosteric signal transmission between the two pockets. In fact, experimental studies have demonstrated that the conformational flexibility of Trp 246, also known as “toggle switch”, is key in guiding the activation/deactivation of the receptor. Pocketron also identified selected frames of the MD trajectory in which the two pockets are merged into a single entity. This led to the formation of a small channel through which the Na⁺ ion could access the cytosol from the extracellular space, as experimentally reported.

Merging and splitting events have been also shown in our MD simulations of the Abl kinase. Here, we focused chiefly on the orthosteric ATP pocket, comparing the results obtained through the analysis of the catalytically active DFG-in and catalytically inactive DFG-out kinase domain (KD_{in} and KD_{out}, respectively). Together with the ATP pocket, Pocketron also revealed the presence of two subpockets. One is an allosteric pocket often exploited by type II kinase inhibitors, while the other is in between the P-loop and the β3-αC-helix loop -present only in KD_{out} (see Figure 4.3). Notably, this region has been targeted by a new class of MEK1 and 2 kinase inhibitors and also accommodated the piperazine-phenyl-pyrimidine moiety of SCH772984 in both ERK1 and ERK2 kinases. This suggests that this pocket can be exploited for designing new compounds that target selectively the inactive (i.e. DFG-out) Abl form. Moreover, our analysis evidenced that split and merge events among ATP and the aforementioned subpockets are more frequent in the inactive KD_{out} than in the active KD_{in} form. This intensive cross-talk reflects an elevated plasticity of the ATP binding site in the case of KD_{out}, as also confirmed

previous computational studies.^{207, 129} As a consequence, the augmented conformational flexibility of the ATP binding site mirrored a diminished catalytic activity. Therefore, through the merge/split analysis, we demonstrated that our tool can capture even subtle protein conformational changes, providing useful insights into the comprehension of protein function.

Lastly, we tested the ability of Pocketron to identify communication networks of pockets spanned across the overall KD of Abl. Interestingly, we found a long-range cross talk between the two distant orthosteric ATP and the allosteric myristate pockets. This cross talk has been largely reported in the literature and represents one among the several allosteric mechanisms that modulate the kinase function. According to the resulting pocket communication network, the allosteric signal passes through a set of pockets located in proximity of α G and α I-helix of the C-lobe, which is a region that can be target by allosteric kinase inhibitors, as suggested by a recent computational study (Figure 4.4). This region, therefore, can be considered crucial to mediate the allosteric signal transmission between the allosteric pocket and the ATP binding site. Indeed, both HX MS experiments and MD simulations indicated that both α G and α I-helix are involved in the allosteric signal propagation. According to our results, the integrity of the network is essential for a proper kinase functionality. In fact, Pocketron detected a perturbation of the pocket communication network only in the dysregulated T315I Abl form, that occur at the level of α G and α I-helix of the C-lobe (Figure 4.4). Therefore, the T315I point mutation induces a structural rearrangement of the pocket communication network, which might explain the deregulation of the T315I Abl form.

Hence, these findings suggest that the interaction network among superficial pockets can be seen as an external reflection of conformational rearrangements occurring also inside the protein²¹² and Pocketron can be considered a new computational method for an effective analysis of allosteric signal propagation.

CONCLUSIONS

In this paper, we present a novel algorithm, implemented in the new software Pocketron, which permits an automated detection and detailed analyses of protein pockets and their spatiotemporal behavior along extended MD simulations. Notably,

this algorithm is able to characterize the history of each pocket, reporting topological events as merge and split, as well as temporal information on their formation and disappearance. Moreover, for each detected pocket, Pocketron returns its persistence and volume along time, listing all atoms and residues that define each pocket. Another key feature of Pocketron is its capability to identify communication networks among near and far pockets and to retrieve allosteric signal propagation pathways.

All these features have been validated here by analyzing extended MD trajectories of selected pharmaceutical relevant biomolecular systems. The results also exemplify how this new algorithm could be successfully employed in the field of drug discovery to identify and characterize new hot spot, transient pockets, and communication network for signaling propagation, without any a priori information on their existence.

THEORY

Outline of the method. In Pocketron, we identify surface pockets and track their evolution along a molecular dynamics (MD) trajectory. A first algorithm is devoted to the detection of pockets on the protein surface, with particular attention to those that could be binding or druggable sites. It has been implemented in the 0.7 version of NanoShaper (freely available at www.electrostaticszone.eu). This algorithm performs such task for every given frame of an MD trajectory. The execution time for the detection of binding pockets in a single frame of a protein of ~2000 atoms is in the range of 2-6 seconds. A second algorithm is dedicated to the characterization of pocket dynamics along an MD trajectory. This new algorithm implements a unique and dynamics-consistent identifier of each atom of every pocket. In this way, the spatiotemporal evolution of each pocket is monitored over the entire MD trajectory. Both algorithms are described in the following sections, while Pocketron is implemented in the BiKiLifeSciences software.

Static pocket detection. The pocket detection algorithm in Pocketron is based on the Solvent Excluded Surface (SES), or Connolly-Richards surface, which is defined as that obtained by rolling a spherical probe over the Van der Waals surface of the molecular system, described as the union of atomic spheres. We computed the SES analytically via the Alpha Shapes theory.¹⁷⁹ Then, Pocketron identifies surface

pockets by making the volumetric difference of the regions enclosed by the SESs obtained with two different probe radii. The smaller rolling spherical probe has a radius of 1.4 Å, which corresponds to the spherical approximation of a water molecule (Figure 4.5). Conversely, the larger rolling spherical probe has a default radius of 3Å (Figure 4.5). We found that these specific size values, which can however be modified at will, together with a subsequent filter selecting only pockets with a volume of at least 11.5 Å³ (~3 water molecules), provide a reasonable identification of pockets and potential binding sites. Thus, Pocketron analyzes the entire protein surface, enumerating the identified pockets and storing their calculated volume and list of contributing atoms-per-pocket. Moreover, Pocketron also provides a grid-based representation of each identified pocket.

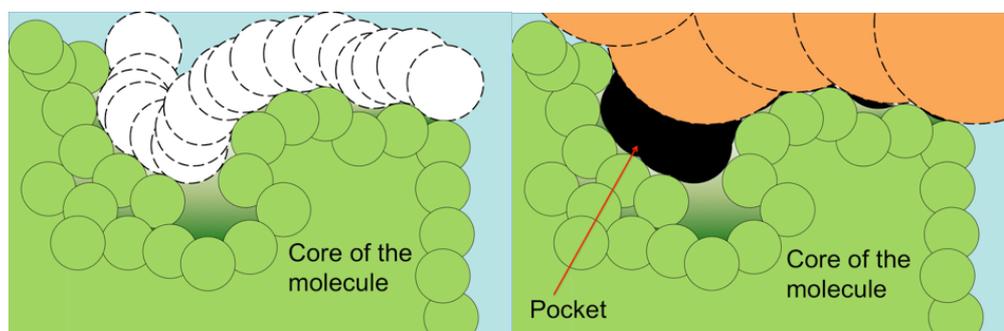


Figure 4.5. Solvent Excluded Surface based pocket definition. The pocket definition here follows the Richards-Lee definition of the solvent excluded surface, that is the surface obtained by rolling a probe over the atomic system as shown in the left panel. In our case, the pockets are identified by performing the volumetric difference between the regions enclosed by the SESs obtained with a larger probe, (right panel, in orange), and a smaller one. In our case the radii values are 3Å and 1.4 Å, respectively. The resulting regions are then duly polished, in order to eliminate spurious ones.

Dynamical pocket tracking algorithm. The atoms forming each pocket are the fundamental elements of our new algorithm for monitoring pocket dynamics in that they represent the actual tracked elements. In fact, the algorithm tracks all the atom-based events occurring along an MD trajectory, without the need of any structural alignment or prior knowledge concerning the region where to search for pockets. In this way, exchange of atoms among adjacent pockets is tracked, indicating cross-talk events between them. By ‘atom exchange’ is meant the fact that the same atom is found to belong to different pockets in different times of the simulation. This

analysis considers as significant so-called “merging” and “splitting” events, which are calculated via simple operations on the sets of atoms that form each pocket. More specifically, a “merge” event occurs when atoms that belonged, in a previous frame, to separated pockets, in the current frame belong to a sole pocket. Symmetrically, a “split” event is when atoms of a previously single pocket get separated and form two or more distinct new pockets. To carefully guarantee consistency in pocket tracking, Pocketron initially stores the list of pockets found in the first frame, as defined by their constituent atoms, and assigns a unique ID to each stored pocket. Then, for every additional frame, the current set of pockets is computed and compared with the stored ones. Only in case of mismatch with respect to all of the stored pockets, the current pocket is stored as a new entry in the stored pocket’s list and assigned a new ID. To measure the degree of similarity between two pockets, Pocketron uses the Jaccard index. For example, let A and B be the sets of atoms belonging to two pockets. Then, the Jaccard index between them is defined as:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

where the modulus symbol indicates the cardinality of the atom sets. If A and B are identical, i.e. have the same constituent atoms, the index reaches its maximum value of 1. In case only a fraction of atoms is shared between A and B, the index provides the number of matching atoms (numerator) divided by cardinality of the union of the atom sets. In case there is no matching, the index is null.

Pocket identification is therefore performed through the similarity matrix \mathbf{M} . Each entry $\mathbf{M}(i,k)$ is the Jaccard index between atom constituting the i -th stored pocket and those of the k -th pocket detected in the current frame. At every frame, each pocket is compared to all of the already stored ones. If $\mathbf{M}(i,k) = 0 \forall i$, then the k -th pocket is considered as a new one and is stored. Otherwise, the k -th pocket will be connected to the one, among the stored ones, which maximizes the Jaccard index. In mathematical terms said p_i the i -th stored pocket and $\widehat{p}_k(t)$ the k -th detected pocket on the frame at time t the \mathbf{M} matrix is therefore the following:

$$\mathbf{M} \triangleq \begin{pmatrix} J(p_1, \widehat{p}_1(t)) & \cdots & J(p_1, \widehat{p}_K(t)) \\ \vdots & \ddots & \vdots \\ J(p_N, \widehat{p}_1(t)) & \cdots & J(p_N, \widehat{p}_K(t)) \end{pmatrix} \quad (2)$$

where K is the total number of detected pockets in the current frame and N is the total number of stored pockets so far. As already mentioned, the pocket ID assignment is done via finding the maximum entry for every column of \mathbf{M} :

$$w_k = \underset{i=1..N}{\operatorname{argmax}} \mathbf{M}(i, k); \forall k = 1..K \quad (3)$$

where the index w_k is the index corresponding to the stored pocket to which k is assigned to. If $\mathbf{M}(w_k, k) = 0$, a new entry, with its own new ID, is created and stored in the pocket list.

Similarly, to detect merging and splitting events, Pocketron builds the matrix \mathbf{F} , so as that $\mathbf{F}(k, j)$ is the Jaccard index of the k -th pocket at instant t and the j -th pocket found in the previous frame at instant $t-1$. Then the \mathbf{F} matrix is:

$$\mathbf{F} \triangleq \begin{pmatrix} J(\widehat{p}_1(t), \widehat{p}_1(t-1)) & \cdots & J(\widehat{p}_1(t), \widehat{p}_j(t-1)) \\ \vdots & \ddots & \vdots \\ J(\widehat{p}_K(t), \widehat{p}_1(t-1)) & \cdots & J(\widehat{p}_K(t), \widehat{p}_j(t-1)) \end{pmatrix} \quad (4)$$

where now J is the total number of pockets at the $t-1$ instant, (i.e. the previous frame). Traversing this matrix allows to easily detect merge and split events. Indeed, termed “nz” the operator that returns all the non-zero entries of a row/column vector and termed “(i,:)” and “(:,i)” the operators which return all the elements of the i -th row/column, respectively, we have that:

$$\begin{aligned} \mathbf{Merge}(\widehat{p}_k(t)) &= \operatorname{arg\,nz}(\mathbf{F}(k, :)) \\ \mathbf{Split}(\widehat{p}_j(t-1)) &= \operatorname{arg\,nz}(\mathbf{F}(:, j)) \end{aligned} \quad (5)$$

where $\mathbf{Merge}(\widehat{p}_k(t))$ is the set containing all the indexes of the pockets detected at the $(t-1)$ -th frame that shared some atoms with pocket k . If this set is empty, then no merge event involved $\widehat{p}_k(t)$. Symmetrically, $\mathbf{Split}(\widehat{p}_j(t-1))$, is the set containing all the indexes of the pockets detected at the t -th frame that shared some atoms with pocket labeled as j at the frame $(t-1)$ -th. If this set is empty, then no split event involved $\widehat{p}_j(t-1)$.

As a technical note, we observe that the entries of the \mathbf{F} matrix directly concern only the pockets detected at the t and $t-1$ instants. However, once the merging and

splitting events are detected, to connect these pockets to the stored ones and to make the tracking consistent, once again the \mathbf{M} matrix is used as previously described.

Detection of a signal transmission network. Given the mentioned tracking capability along time, the volume and surface area of all pockets are stored, at every frame. Thus, at the end of the analysis, Pocketron has a set of atoms characterizing each pocket and the full history of the volume, area, splitting and merging events, for all the unique pockets detected along the MD run. During the tracking, atoms are used to identify the best matching ID, nevertheless the instantaneous information of the residues that at instant t characterize a given pocket is stored too. This allows to compute the probability – expressed as a relative frequency – of each residue to be part of a given pocket. In addition, for each pocket, all of the merging and splitting events with other pockets are stored. Indeed, one of the final result of the tracking process is a square matrix \mathbf{N}_M , the (i,j) -th entry of which represents the merging probability between two pockets expressed as the number of times they merge over the total number of frames. Similarly, the square matrix \mathbf{N}_S expresses the corresponding statistics for the splitting events. Empirically, we found that \mathbf{N}_S and \mathbf{N}_M corresponding entries very often coincide in magnitude, indicating that splitting and merge events often occur at the same time. For this reason, we keep as aggregate statistics the matrix \mathbf{N} that for each entry collects the max value of the corresponding entries in \mathbf{N}_M and \mathbf{N}_S . The non-trivial nature of the information stored in \mathbf{N} calls for effective graphical methods to represent binding pocket dynamics, also considering that \mathbf{N} is highly sparse. Pocketron translates the merging and splitting frequency matrix \mathbf{N} into a 3-dimensional network graph, where the nodes of the graph are pockets, while the edges indicate the communication between two pockets. In detail, the position of each node (i.e., a pocket) is the center of mass of the atoms that form that pocket. The color of the pocket indicates its persistency over the simulated time. The more persistent is the pocket, the lighter is its color. The thickness of the edge is directly proportional to the frequency of merging and splitting events between the two connected pockets. That is, the higher is the communication between the two pockets, the thicker is the edge in between the two pockets.

METHODS

Structural models. In the present work, we employed Pocketron for analyzing the MD simulations of three different systems, namely Purine Nucleoside Phosphorylase (PNP), Adenosinic Receptor (A_{2A}) and Abelson (Abl) kinase. The homotrimeric construct of PNP was modelled using 3K8O X-ray structures and was simulated in presence of nine DADME-ImmH ligands and phosphate ions, retrieved from 1RSZ and 1RR6 PDB structures, respectively. A_{2A} receptor was built using chiefly 3UZC X-ray structures, while 4EIIY X-ray structure was used as template for the missing ECL2 in 3UZC. The apo protein was embedded in POPC membrane. For Abl we built 5 different model systems. Two of them include the apo wild-type KD alone both in the DFG-in (KD_{in}) and in DFG-out conformation (KD_{out}). Two consist of the KD mutated at the gatekeeper residue T315 – located at the ATP binding site – with an isoleucine residue. These mutated forms are either in the apo ($T315I-KD_{in}$) or in complex with myristate ($Myr/T315I-KD_{in}$). The last system is the wild type KD in complex with myristate (Myr/KD_{in}). The KD_{out} system was built starting from 1OPL chain B X-ray structure after the removal of the SH2 domain. The remaining systems were modelled starting from the 2F4J X-ray structure. For further details see Supporting Information.

MD simulations. PNP was parameterized using Amber ff99SB-ildn force field, while General Amber Force Field (GAFF) was employed to parameterize the ligands and the ions. Ligand partial charges were fitted using the RESP procedure via Antechamber. The system was immersed in TIP3P water box and consisted of ~100,000 atoms. After 350 ns of equilibration which led the system to a temperature of 300 K and to a pressure of 1 bar, we run the production in NVT ensemble. We collected ~710 ns of simulation stored in ~1800 frames for Pocketron analysis.

A_{2A} was parameterized using the Amber ff99SB-ildn force field. The protein and the POPC membrane were immersed in TIP3P water box reaching a total of 65802 atoms. Also, here the system was first equilibrated to reach a temperature of 300 K and a pressure of 1 bar, while the production was run for 100 ns in NPT ensemble. We extracted from the trajectory ~1000 frames for Pocketron analysis. The five Abl kinase systems were parameterized using the Amber ff99SB force field. For $Myr/T315I-KD_{in}$ and Myr/KD_{in} models the myristate was parameterized using

General Amber Force Field (GAFF) after computing the point charges at HF/6-31G* level theory. The systems were solvated in TIP3P water and consisted of ~45000 atoms. After 5 ns of equilibration the systems reached a temperature of 300 K and a pressure of 1 bar, we run the production in NPT ensemble for a simulation time that spans from ~0.8 μ s to ~2.5 μ s. Here, Pocketron was used to analyze from ~16,000 to ~50,000 frames. See SI for more details.

Pocketron parameters. For each Pocketron analysis, we set the small and the large probe radii respectively to 1.4 and 3 Å. To avoid unnecessary noise due to the detection of very small pockets, we set the minimal volume of a detectable pocket to the equivalent of 5 water molecules for PNP and A_{2A} and 4 water molecules for Abl. Finally, we want to stress that for tracking purposes, we removed from each system trajectory all the ligands, the ions, the membrane, water molecules, taking into account only the protein.

Supporting Information

Details on systems preparation, figures and tables reported in the main body are available at Appendix B.

Chapter 5

Structural characterization of BTK kinase via integrative modeling techniques

INTRODUCTION

Bruton's tyrosin kinase (Btk) is a non-receptor kinase that belongs to Tec family, chiefly expressed in hematopoietic cells, as B cells, platelets,²¹³ neutrophils²¹⁴ and macrophages.²¹⁵ Btk has a prominent role in B-cell antigen receptor (BCR) signaling, fostering the development of B-cells.²¹⁶ Mutations in Btk gene often results in the lack of Btk protein or in its impaired function, which is the main cause of the overcoming of primary immunodeficiency X-linked agammaglobulinemia (XLA).²¹⁷ Such a disease is characterized by the block of B cell development, leading to the lack of antibodies and causing, in turn, an increased susceptibility to infections. Therefore, given its important role in B-cell development and function, Btk is nowadays an attractive drug target for the treatment of B-cell malignancies and autoimmune disease.^{218, 219}

Btk has a modular structure similar to that of Src family kinases, containing the SH3, the SH2 and the catalytic Kinase Domain (KD). In addition, Btk is also endowed of the Pleckstrin Homology (PH) domain, which allows the binding with the phosphatidylinositol (3,4,5)-trisphosphate (PIP₃) in membrane, and the Tec Homology (TH) domain, which contains a zinc finger motif, which is important for protein stability and function.^{220, 221} Recent biophysical and structural studies revealed that, when in the inactive (or auto-inhibited) form, Btk assumes a closed and compact conformation.²²² Here, the SH2 and the SH3 domains are bound the lateral part of KD as also reported for Src family kinases and Abl kinase. The PH and TH domains, instead, are located on the back of the SH3-SH2-KD complex. When active, Btk seems to adopt an elongated conformation, as reported by previous small-angle X-ray scattering (SAXS) measurements of full-length Btk.²²³ The activation of Btk is initiated by its interaction with PIP₃ at the membrane. The membrane recruitment triggers the trans-auto phosphorylation both at Tyr551, which is the canonical phosphorylation site at the activation loop (A-loop) and at Tyr233 in the SH3 domain.^{224, 225} Phosphorylation, in turn, releases the chemical constraints that lock Btk in the closed conformation, allowing the formation of the elongated configuration of Btk, as also observed in several other kinases.²²³ Recently, the elongated active conformation has been also reported in Abl kinase through X-ray crystallography.⁷⁰ These X-ray structures show that the SH2 domain binds the N-lobe of KD. Interestingly, biochemical and computational studies have found that the

binding of the SH2 allosterically increases both the catalytic activity and the phosphorylation of the Tyr412 at the A-loop, a required step for full kinase activity.^{207, 129, 128} These findings demonstrated that the SH2 domain in Abl not only has a role in mediating the protein-protein interaction, but also modulates of kinase activity. In fact, the disruption of the SH2-KD interaction is considered a viable strategy to inhibit the Abl function.⁷⁶

As mentioned before, likewise to Abl, Btk also contains the SH2 domain and adopts an elongated conformation when active. Therefore, it is reasonable to hypothesize that also in Btk, the SH2 might both modulate allosterically the kinase activity and favor the phosphorylation of the A-loop at Tyr551. However, unlike Abl, detailed structural insights on the SH2-KD interaction of Btk are missing. Notably, this information is essential to unveil the molecular mechanisms responsible for the allosteric signal propagation across the SH2-KD complex. In the present work, we integrated molecular modeling and biophysical techniques to provide structural insights on the SH2-KD interaction in Btk. To do so, we firstly performed enhanced sampling molecular dynamics (MD) simulations to retrieve an ensemble of SH2-KD conformations that were subsequently compared with SAXS data. As preliminary result, this procedure enabled us to select SH2-KD complexes that better fit the experimental data.

RESULTS AND DISCUSSION

In the present work, we coupled computational studies, based on enhanced sampling MD simulations, with structural and biochemical studies to provide structural insights on the formation of the activated SH2-KD complex in Btk kinase. In particular, through MD simulations, we performed a dynamic domain-domain docking in order to probe the interactions that favor the SH2-KD binding. Then, we retrieved an ensemble of relevant SH2-KD complexes from MD trajectories that were integrated to SAXS data in order to propose a model of the active SH2-KD complex.

Prediction of SH2 anchor point residues. Firstly, we focused our attention on the identification of key residues of the SH2 that might have a role in mediating the interaction with KD. Toward this aim, we performed a bioinformatics-based

analysis, by means of the whispy web server. This allowed predicting surface residues of SH2 domain that are more likely to be involved in the binding with KD. As a result, whispy predicted 12 residues as putative interaction points of the SH2 domain (red residues in Figure 5.1) and 25 residues as putative surrounding residues (green residues in Figure 5.1). These residues are clustered in two distinct sites of the SH2 domain, which are referred to as S1 and S2. S1 corresponds to surface binding site nearby the phospho-peptide binding pocket, while S2 is in between the α B helix and the SH2-KD linker (Figure 5.1). Since S1 is already involved in the binding with the peptide substrate,²²⁶ it is unlikely that it interacts also with KD. Therefore, we discarded S1 from subsequent analysis, focusing our attention only on S2.

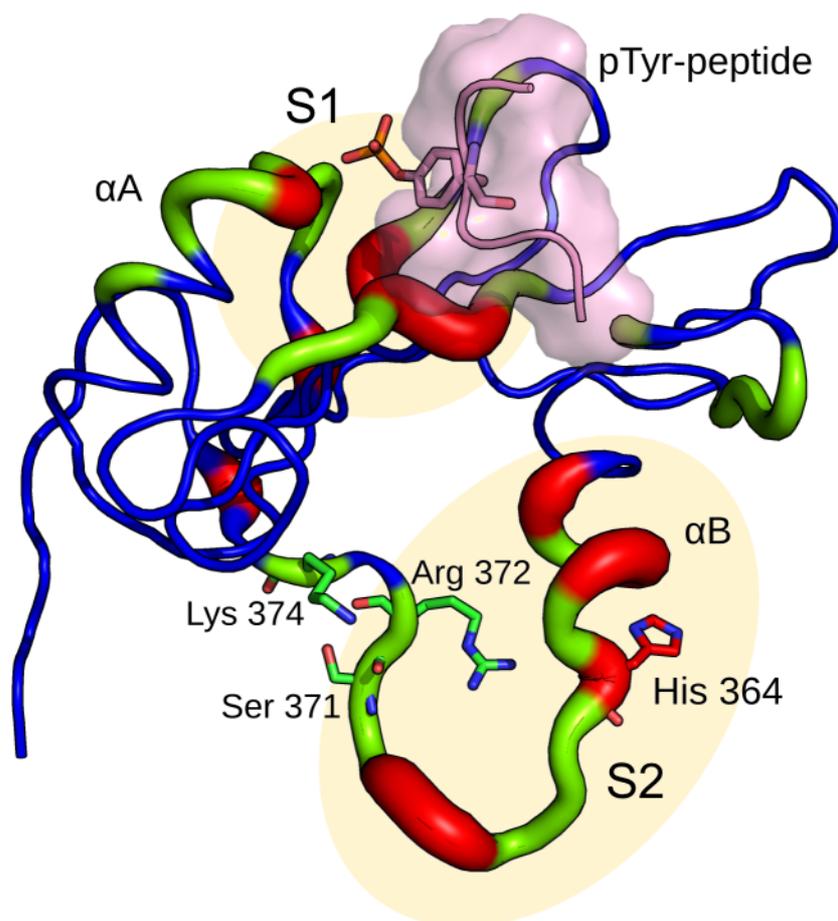


Figure 5.1. SH2 domain of Btk kinase (pdb code: 2GE9). The portions in red represent the residues that whispy predicted as putative anchor points for the PPI, while the residues in green are the predicted surrounding spots. As illustrative scope the peptide is depicted in pink. It has been retrieved by Itk NMR structure with the code 2ETZ.

To further corroborate our hypothesis, according to which S2 is involved in the interaction with KD, our performed mutational studies on SH2 domain. In order to select a set of SH2 residues to mutate, we crossed the information obtained by whispy prediction and known point mutations that are found in XLA-patients.²²⁷ As a result, we selected the following Btk mutations: His364Asp, Ser371Pro, Arg372Gly, Lys374Asn. To assess whether a point mutation in SH2 domain either favors or hampers the binding with KD we set up a biochemical assay as suggested in Grebien et al.⁷⁶ and Lamontanara et al.¹²⁸ Briefly, in this assay we measured the phosphorylation of the A-loop at the Tyr551, which is known to be increased upon binding of SH2 domain with KD. Therefore, mutations that prevent the SH2-KD binding should return a decreased A-loop phosphorylation compared to the wild type. As a result, in all the Btk mutants the phosphorylation at Tyr551 is diminished respect the wild-type construct, which conserved its basal phosphorylation rate (Figure 5.2). Hence, these results indicated that His364, Ser371, Arg372 and Lys374 might be important for SH2 binding, having a role in mediating the interaction with KD.

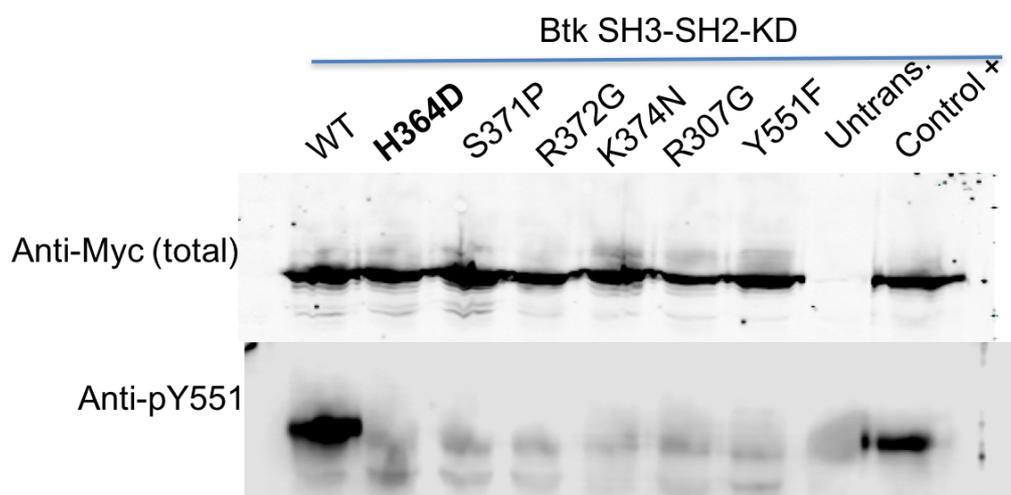


Figure 5.2. The indicated SH3-SH2-KD constructs both in wild-type and mutated form were analyzed by immunoblotting with an anti-pY551 antibody.

Role of the linker in SH2-KD complex formation. At this point, we focused our attention on the SH2-KD linker with the aim of understanding its role in the formation of SH2-KD complex in Btk kinase. To do so, we performed a ~200 ns long aMD simulation of SH2-KD system in absence of the linker (see Methods). We

monitored the binding events of the two domains during the simulations, computing the distance between the center of mass (COM) of the SH2 domain and the COM of the N-lobe of KD (i.e. dI). In the first ~ 80 ns of simulations, dI oscillates between ~ 40 and ~ 90 Å, indicating that the two domains approached each other even though binding events have not been detected (Figure 5.3). From ~ 80 ns until the end of the trajectory, the distance between SH2 and KD increases reaching a value of ~ 150 Å, indicating that the SH2 moved away from the KD (Figure 5.3). These results indicate that in absence of the SH2-KD linker the two domains do not bind in our simulations.

To confirm our findings, our collaborators performed size exclusion chromatography and isothermal titration calorimetry (ITC) experiments in order to evaluate the interaction between the SH2 domain and KD (see Methods). Notably, both experiments revealed that the two domains do not interact to each other, further corroborating the results obtained through aMD simulations. Therefore, the interaction between the SH2 and KD is impaired without the SH2-KD linker, which, in turn, might be a key structural element that ensure a proficient domain-domain interaction.

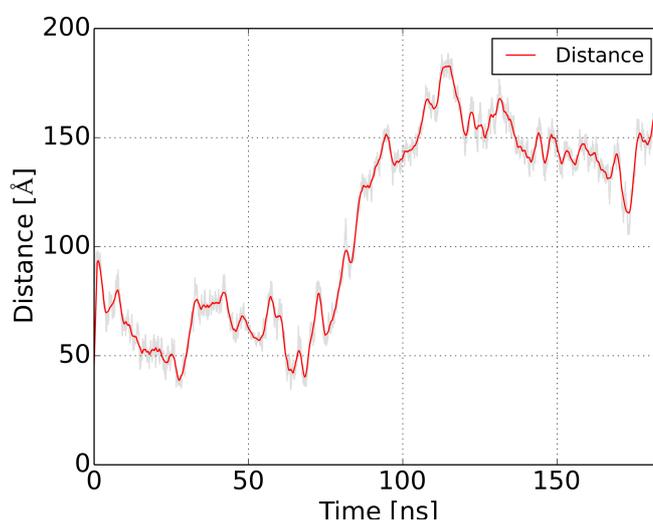


Figure 5.3. Time evolution distance between the COM of SH2 and KD in SH2-KD system.

Investigating the binding events through aMD simulations. Here, we performed aMD simulations of Btk SH2-KD model in presence of SH2-KD linker

(i.e. SH2-KD-Linker, see Methods) in order to study domain-domain binding dynamics and to retrieve potential structures of the SH2-KD complex. To increase the sampling of the phase space, we run 11 independent replicates that last ~200 ns each, for a total of ~2.2 μ s of aMD simulations. After visiting several transient and unbounded states, a binding event occurred in all replicates.

Firstly, we focused our attention on the role that His364, Ser371, Arg372 and Lys374 have in mediating the binding between the SH2 domain and KD. To do so, we computed the number of contacts between the aforementioned residues and both the SH2-KD linker and N-lobe of KD during the simulations (see Methods). As a result, His364, Ser371 and Arg372 often interact with the SH2-KD linker, having a frequency of contacts from 12 to 20% (see orange and red spots in Figure 5.4). This suggests that during our simulations, the linker adopts a secondary structure that facilitate the interaction with the SH2 domain, confirming our previous hypothesis according to which the linker plays a structural role in favoring the SH2-KD binding. Besides, His364, Ser371, Arg372 and Lys374 also interact with the N-lobe. In particular, His364 preferably interact with a region in proximity of the β 3- α C loop (see Figure 5.4). Instead, Ser371, Arg372 and Lys374 rather bind the back of KD in between the β 2- β 3 loop and the α C helix. (see Figure 5.4). These results indicate that His364, Ser371, Arg372 and Lys374 favor the formation of SH2-KD complex, because they interact directly with precise portions of both the linker and N-lobe of KD. Notably, these findings are in line with previous bioinformatics and biochemical studies, further confirming the role of His364, Ser371, Arg372 and Lys374 in stabilizing the bound form of SH2-KD.

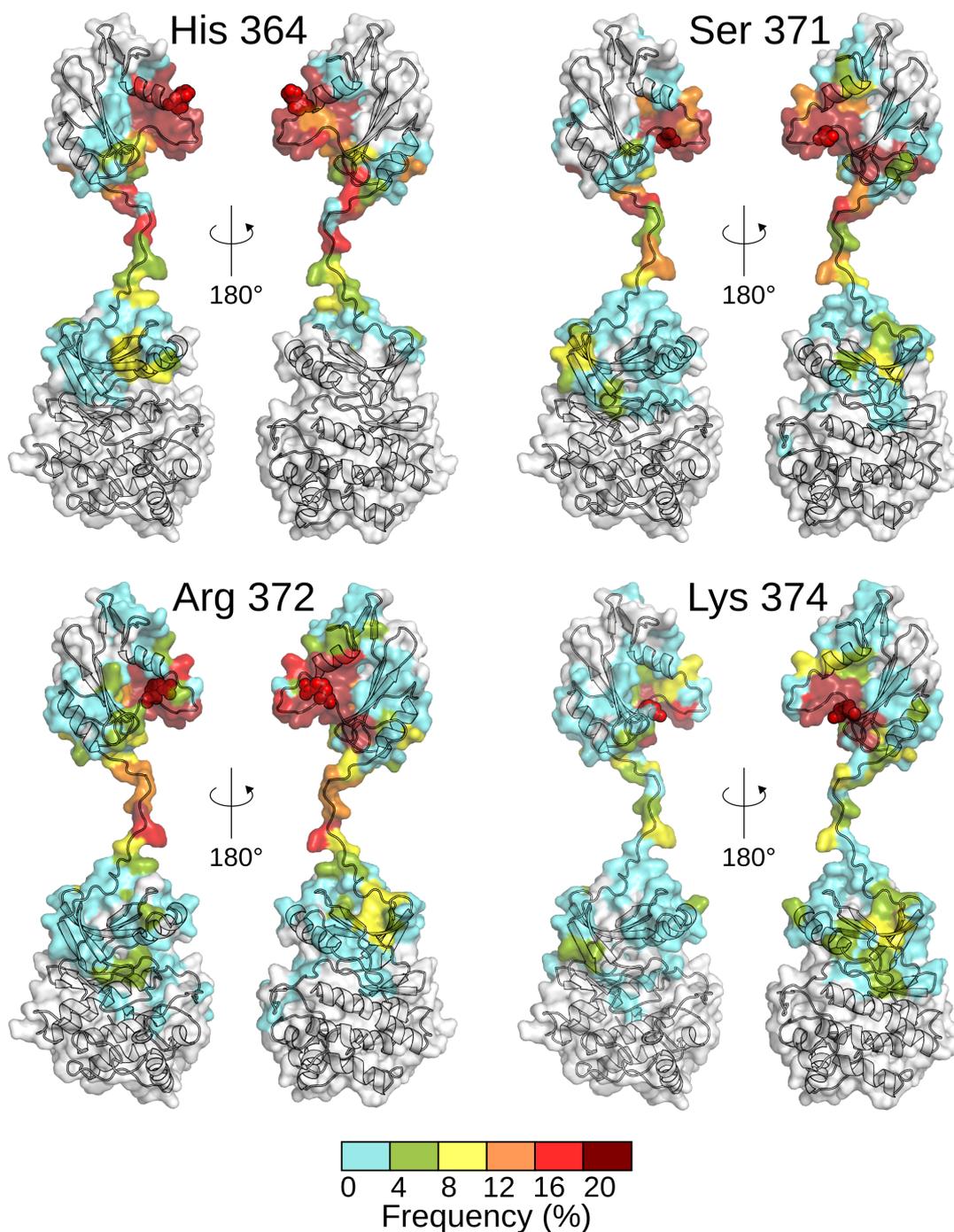


Figure 5.4. Representation of contact frequency of His364, Ser371, Arg372 and Lys374 mapped onto SH2-KD complex (starting frame of all MD trajectories). The color code represents the frequency of contacts along the simulations.

Comparison of aMD simulations with SAXS data. At this point, we performed a cluster analysis to retrieve the most populated configurations of the SH2-KD complex sampled during the MD simulations and compare them with the SAXS data generated by our collaborators. The final goal is to retrieve SH2-KD configurations

having a theoretical SAXS scattering curve similar to the experimental one and fit the complexes in the SAXS envelop. Overall this procedure allows us to determine the most probable SH2-KD complexes that better fit the experimental data.

The cluster analysis returned nine main clusters which are populated from ~1 to ~30% of the simulations time (see Figure 5.5). From each cluster, we extracted the centroid structure, which are depicted together in Figure 5.5. From here, it can be noted that the SH2 domain assumed three main spatial orientations respect to KD. In particular, it can be found on the “central-right side” of KD, as cluster 1 (C1) and C2, on the “back side” of KD, as C3, C7 and C9 and finally on the “top side” of KD as C4, C5, C6 and C8. Interestingly, the top side orientation of the SH2 domain is the most populated one (47.2% against 17.6% and 35.2% of the central-right side and back side, respectively), suggesting that the elongated complex is the most probable for Btk kinase. It is worth pointing out that some clusters have been submitted to a re-cluster analysis in order to remove the noise and define sub-configurations of the main cluster. For example, C3 has been divided in six sub-clusters (i.e. C3-1 to C3-6). Through this procedure we obtained a total of 28 clusters.

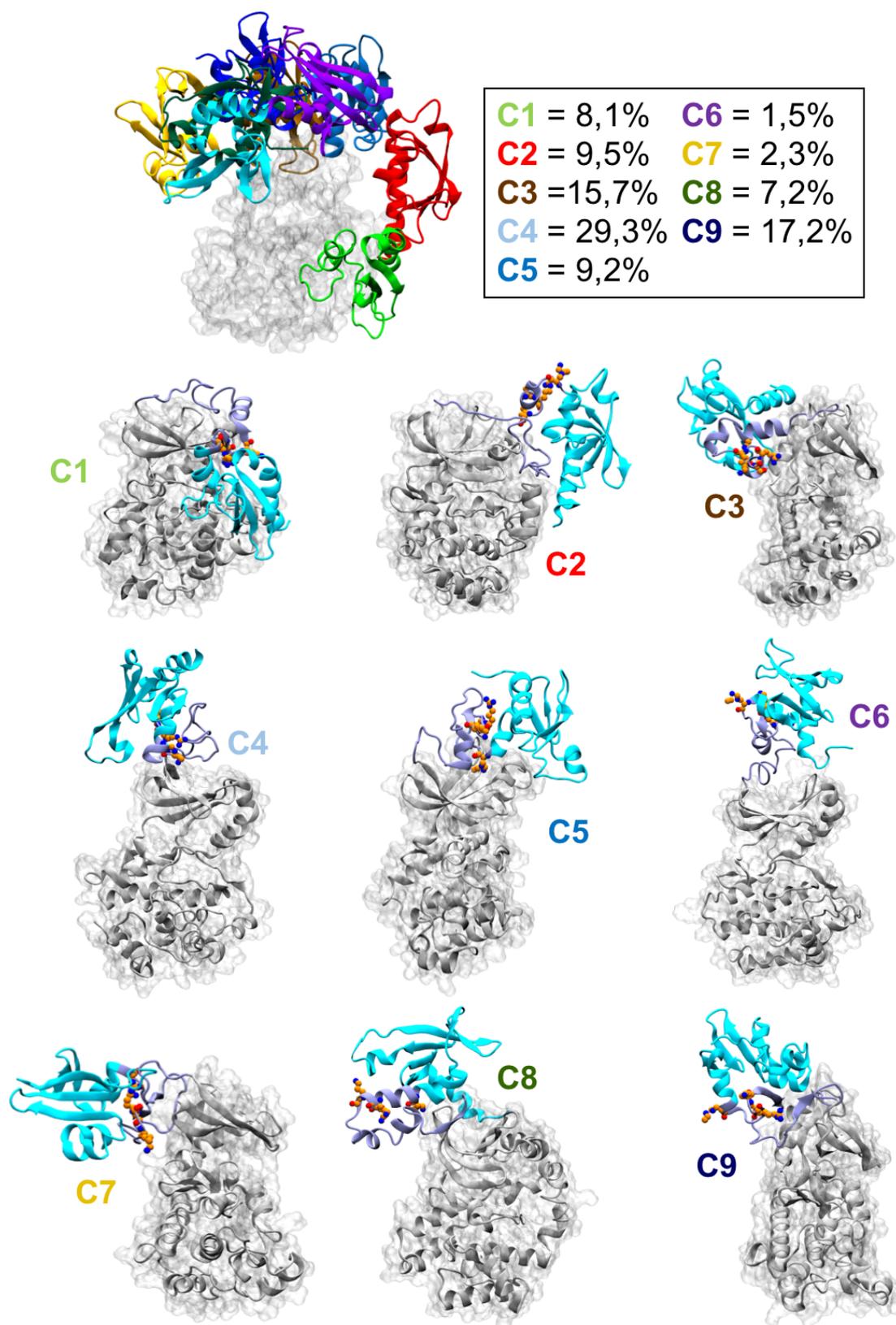


Figure 5.5. Representation of the centroids of the nine clusters obtained from aMD ensemble. On the top the nine centroids are superimposed in order to distinguish the different SH2 spatial orientation. Also, the population of each cluster, in terms of percentage over the entire simulation, is reported. On the bottom, the nine centroids

are depicted separately. The SH2 is colored in cyan, the KD in grey, while the SH2-KD linker in violet. The residues His364, Ser371, Arg372 and Lys374, discussed in the main body, are represented in orange.

At this point, we employed the software Crysol to evaluate the theoretical scattering curves from the 28 cluster centroids and fit them to the experimental SAXS scattering curves. The quality of the fit is a function of the computed χ^2 value. The lower the value the better is the fitting with the experimental curve. From this procedure, four clusters, namely C4-1, C4-2, C5-2 and C6, resulted to better fit the experimental scattering curve, having a χ^2 value of 2.14, 1.65, 1.94 and 2.50, respectively (see Figure 5.6 A and Table 5.1). Remarkably, all these clusters have an elongated conformation where the SH2 domain is on the top side of KD. These data are in agreement with the results obtained from cluster analysis, indicating that the most populated configurations of our aMD simulations are also those that better fit the experimental data. In addition, also the computed radii of gyration $R(g)$ of C4-1, C4-2, C5-2 and C6 are in agreement with the experimental one (30.7, 31.7, 30.9 and 31.2 Å of C4-1, C4-2, C5-2, C6, respectively, against 34.3 Å of experimental one). Subsequently, by means of Situs package, we performed an automated fitting procedure of C4-1, C4-2, C5-2 and C6 3D structures with the SAXS envelop. The docked structures are depicted in Figure 5.6 B.

Table 5.1. List of theoretical chi-square values (χ^2) and radii of gyration ($R(g)$) of 28 cluster centroids computed by means of Crysol.

	χ^2	$R(g)$		χ^2	$R(g)$
C1	25.79	24.30	C5-2	1.94	30.98
C2	12.89	26.40	C5-3	5.96	28.79
C3-1	8.14	27.45	C5-4	5.15	28.88
C3-2	11.65	26.57	C5-5	4.42	29.54
C3-3	7.94	27.52	C6	2.50	31.23
C3-4	10.64	27.11	C7	9.52	27.17
C3-5	7.96	27.76	C8-1	9.90	27.22
C3-6	9.32	27.19	C8-2	9.28	27.67
C4-1	2.14	30.71	C9-1	4.81	29.21

C4-2	1.65	31.79	C9-2	5.62	28.82
C4-3	5.49	28.68	C9-3	11.87	26.73
C4-4	5.13	28.93	C9-4	7.35	28.69
C4-5	5.77	28.43	C9-5	7.21	28.02
C5-1	8.36	27.67	C9-6	7.18	28.22

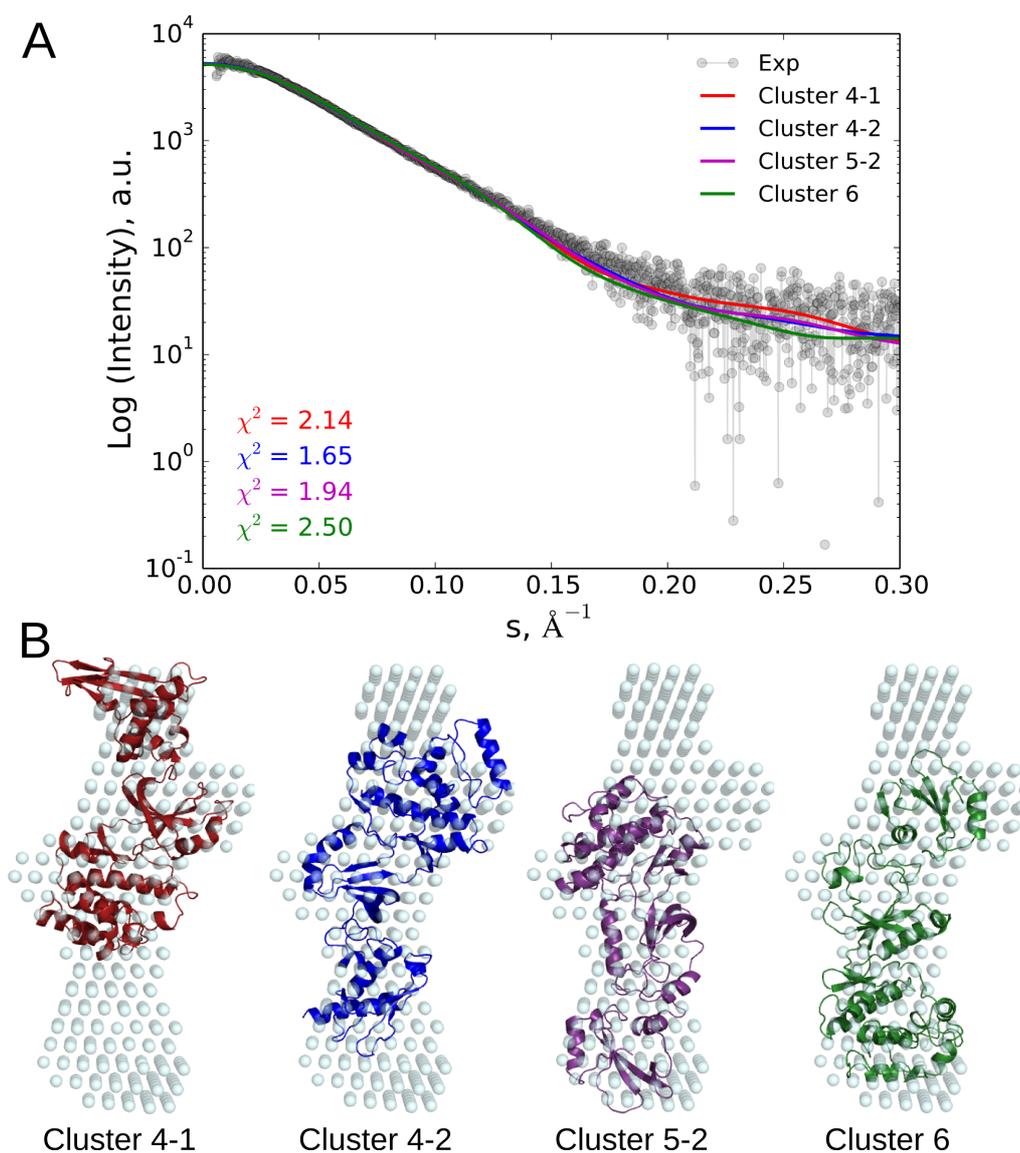


Figure 5.6. A) Fitting of theoretical curves of C4-1, C4-2, C5-2 and C6 with the experimental scattering SAXS curve. B) Each 3D structure of the four centroids are docked into the SAXS envelop (cyan spheres). Only the best scored structures are reported.

Ongoing simulations and experiments.

Preliminary results showed that selected structures retrieved from aMD-generated conformational ensemble have a spatial orientation of both KD and SH2 domain that is in agreement with SAXS experimental data. In particular, theoretical scattering curves of C4-1, C4-2, C5-2 and C6 structures present a low χ^2 value (2.14, 1.65, 1.94 and 2.50, respectively). However, a further refinement of the models is needed to decrease the χ^2 of the four structures. Toward this aim, additional classical MD simulations of C4-1, C4-2, C5-2 and C6 structures are running in order to relax the structures. Selected frames from these 4 MD trajectories will be newly compared with the experimental SAXS scattering curve and only the best complexes will be fitted to the SAXS envelop.

Besides, a new Btk SH2-KD construct has been prepared and will be submitted to new SAXS experiments by our collaborators. Unlike the construct used in the reported SAXS experiments, this new construct is not endowed of an N-terminal tag, which has been found to cause noise during the modeling of the SAXS envelop. In fact, as can be seen in Figure 5.6 B, the resulting map of the tagged-Btk is remarkably larger compared to the space occupied by the SH2-KD complex. This, in turn, negatively affects the docking of our atomistic models with the SAXS envelop.

METHODS

Experimental section

Size Exclusion and ITC. Btk kinase domain (residues 391-659) was expressed and purified. Btk SH2 domain (residues 270-390) was cloned into a pETM30 vector and expressed in *E. coli* during 4 hours at 28°C. Cells were lysed in lysis buffer and purified using Ni-NTA beads (Qiagen). Protein was eluted in buffer containing 50mM Tris pH7.5, 500nM NaCl, 1mM DTT, 300mM imidazole and 5% glycerol and the 6xHis-tag cleaved by TEV protease during overnight dialysis at 4°C in buffer 50mM Tris pH7.5, 150nM NaCl, 1mM DTT and 5% glycerol. After a reverse Ni-NTA purification, the flow-through containing the SH2 domain was further purified by size exclusion chromatography in the same buffer, concentrated and frozen. Intermolecular interactions between the Btk kinase and SH2 domains were first tested by size exclusion chromatography. First, purified proteins were run separately

in a Superdex75 10-300 column (GE Healthcare) to verify their elution pattern. Then, a sample containing Btk kinase domain and SH2 at equimolar ratio was incubated at 4°C during 4 hours and run on the S75 10-300 column to detect any peak shifts in case of interaction. The isothermal titration calorimetry (ITC) assay was performed using the MicroCal™ iTC200 (GE Healthcare). Purified Btk kinase domain and SH2 were dialysed overnight against buffer containing 25mM Tris pH7.5, 40mM NaCl and 1mM TCEP at 4°C. Thermodynamic measurement was done by placing 300 uL of 45uM KD in the sample cell and adding 100ul of hBtk-Sh2 450uM (ratio 1:10) stepwise by 16 microinjections at 25°C. Data was analyzed using the software provided by the fabricant.

Small-angle X-ray scattering (SAXS). A kinase dead (D521N) version of Btk containing the kinase and SH2 domains (residues 270-659) was cloned into a pFAST-BAC-Htb vector and expressed in Sf9 insect cells for 72 hours at 28°C (Bac-To-Bac expression). Cells were lysed in buffer containing 50mM Tris pH7.5, 500nM NaCl, 1mM DTT, 10mM imidazole, 5% glycerol and 10U/ml benzonase. Lysate was centrifuged 35 minutes at 75,000xg and purified using Ni-NTA beads (Qiagen). Protein was eluted in buffer containing 50mM Tris pH7.5, 500nM NaCl, 1mM DTT, 300mM imidazole and 5% glycerol and purified by size exclusion on a Superdex75 16-60 column (GE Healthcare) in buffer containing 50mM Tris pH7.5, 150nM NaCl, 1mM DTT and 5% glycerol. Purified protein was frozen and aliquoted with concentration ranges varying from 0.5 to 2.8mg/ml. Sample purity and molecular mass were verified by mass spectrometry, gel filtration and coomassie staining. Small-angle X-ray scattering (SAXS) measurements were performed at EMBL P12 beamline (DESY, Hamburg). In brief, Btk SH2-Kinase protein samples were measured using an automatic robot²²⁸ at the beam line and data recorded at 10°C using a PILATUS 1M pixel detector (DECTRIS), and data processing performed by automated pipeline. Further analysis of scattering $I(0)$ and the radius of gyration (R_g) were performed in PRIMUS, pair-distance distribution function $P(r)$ evaluated in GNOM, and porod volume and molecular mass estimated. Preliminary ab initio models were computed using DAMMIF²²⁹ and GASBOR²³⁰ for bead and dummy residue models respectively, using data from the range $0.02 < s < 0.15 \text{ \AA}$.

Computational section

System preparation. In this work, we simulated the binding of the SH2 domain on KD of Btk in presence and in absence of the linker sequence. These two systems are referred to as SH2-KD-Linker and SH2-KD, respectively. Both systems were modelled starting from the X-ray structure of KD and the NMR structure of the SH2 domain with the PDB codes 1K2P and 2GE9, respectively. In detail, KD consists of residues from 397 to 654, while SH2 from 270 to 387. The SH2 domain was positioned on the top of KD at a distance of ~ 60 Å from it. The distance is measured considering center of mass of SH2 and N-lobe of KD. In SH2-KD-Linker system, the linker sequence (residues from 388 to 396) was manually according to the FASTA sequence. The models were built by means of Maestro of Schrödinger suite. Hydrogens atom were using tleap of AmberTools14. All the MD simulations were performed in implicit solvent. The protein was parameterized using the ff14Bonlysc, which is ff99SB in combination with the side chain parameters from ff14SB. The GB-Neck2 implicit solvent model was employed in combination of mbondi3 radii (Nguyen, Simmerling 2013, 9, 2020-2034). We chose such set up because was successfully employed in a similar work carried out by Nguyen et al.²³¹

aMD simulations. Firstly, the system was minimized using steepest descent algorithm for 250 cycles and with conjugate gradient for the remaining 250 cycles. The system was heated until 300 K in 500 ps in Langevin dynamics using a collision frequency γ of 1 ps^{-1} . SHAKE algorithm was employed and only bonds involving hydrogens were constrained. All the protein atoms were restrained applying a weight of $10 \text{ kcal/mol} \cdot \text{Å}^2$. In order to better equilibrate the linker, three additional steps MD simulation of 500 ps each were carried out, restraining only the SH2 and KD parts using a weight of 10, 1 and $0.1 \text{ kcal/mol} \cdot \text{Å}^2$. Finally, all the restraints were removed for the production phase which consisted of 5 ns of classical MD simulation.

Accelerated MD (aMD) simulations were performed applying a boost potential to all dihedral angles of the system with the input parameters E_{dihe} and α_{dihe} , where:

$$E_{\text{dihe}} = V_{\text{dihe_avg}} + 4 \cdot \text{solute residues}$$
$$\alpha_{\text{dihe}} = 1/5 \cdot 4 \cdot \text{solute residues}$$

$V_{\text{dihed_avg}}$ is the average dihedral energy obtained in the classical MD simulation.¹⁰⁸ Samples were collected every 0.2 ps. The simulations were run using PMEMD module of Amber14 with GPU acceleration.

Data analysis

Prediction of anchor/interaction points of SH2 domain. Whispy (What Information does Surface Conservation Yields) web server was adopted to predict the residues of the SH2 that might be the anchor points for the interaction with KD. In brief, whispy is based on the assumption that conserved surface residues across homologous proteins have a high propensity to interact with other proteins.²³² As input file of whispy run, we gave a multiple sequence alignment (MSA), computed via M-coffee web server, of 32 tyrosine kinases and the Btk FASTA, as reference structure.

Contacts analysis. Here, we identified the contacts between a set of residues in the SH2, which have been found to be important to mediate the SH2-KD interaction and the KD. The residues are His364, Ser371, Arg372 and Lys374. To do so, we employed an in-house Tcl-based script that, for each frame of the trajectory, recorded the residues of KD around 5 Å of the aforementioned residues. The results were subsequently plotted directly on the KD structure using an in-house python-based script. The analysis was performed after merging all the 11 replicates for a total of ~2.2 μs of aMD simulations.

Clustering analysis. All the trajectories were submitted to a clustering procedure using an in-house software. First, the trajectories were submitted to a PCA analysis in order to dimensionally reduce the data. Only the α carbons were considered for the analysis. The eigenvectors that described the 80 % of the overall variance were retained for the analysis. Subsequently, the trajectory plotted on the relevant eigenvectors was submitted to the clustering procedure by applying a hierarchical scheme. Also here, the analysis was performed after merging all the 11 replicates.

Scattering data analysis. Crysol²³³ was employed to fit the theoretical scattering of the atomic models with the experimental one. Atomic models with the lowest χ^2 value were subsequently docked into the SAXS low-resolution map by means of Situs package.²³⁴

Chapter 6

Drug discovery projects

1. RhoJ

1.1. Introduction

Rho GTPases are proteins belonging to the Ras superfamily that play important role in several cell processes such as cell morphology, motility, proliferation, differentiation and apoptosis.²³⁵ Rho GTPases can be found in active GTP-bound and in the inactive GDP-bound conformation. Importantly, their activity is positively controlled by guanine nucleotide exchange factors and negatively controlled by GTPase-activating proteins. Once loaded with the GTP, the GTPase proteins bind to several substrates, activating a set of downstream effectors.²³⁶ The Rho GTPase family contains 21 members,²³⁷ among which is RhoJ (or TCL) a protein characterized only in 2000.²³⁶ RhoJ shares the 85% and the 78% amino acid similarity with TC10 and Cdc42, respectively. Like TC10 and Cdc42, RhoJ binds the Cdc42/Rac interacting binding domain (CRIB) of PAK and WASP.

A study performed by Ho *et al.* demonstrated that RhoJ is a regulator protein that plays a fundamental role the resistance to chemotherapy in melanoma cells.²³⁸ In particular, RhoJ binds and activates PAK1 kinase in response to DNA damage, which in turn activate several downstream effectors leading to a decreased DNA damage–induced apoptosis. In addition, the same group found that RhoJ is also implicated in melanoma cell migration and invasion *in vitro*, by modulating actin cytoskeletal dynamics in a PAK1-dependent manner.²³⁹ Taken together this results indicated that RhoJ is a viable pharmaceutical target for the treatment of melanoma.

In particular, molecules that disable the RhoJ/PAK1 pathway might both increase the sensitivity of melanoma tumors to DNA damage and inhibit melanoma invasion and metastasis. In this context, we performed a structure-based virtual screening campaign in order to identify compounds able to target RhoJ/PAK1 interaction. As a result, we found six hit compounds active at low micromolar range, which currently are in the lead optimization phase in our labs.

1.2. Computational methods

Protein preparation. Since the lack of X-ray structure of RhoJ protein, we built a homology model to use as receptor for the virtual screening campaign. As template structure, we decided to employ the Cdc42 protein in complex with the CRIB domain of Pak6 (PDB code 2ODB, resolution of 2.4 Å, Figure 6.1). Despite TC10 has a higher sequence similarity with RhoJ than Cdc42 (85% and 78%, respectively), we decided to employ Cdc42 because 2ODB is the only structure complexed with a CRIB domain of PAK kinase. This choice allows building a more reliable model of RhoJ for the study of RhoJ/PAK1 interaction. By means of Prime software implemented in Maestro, we modeled the FASTA sequence of RhoJ on the Cdc42 template. At this point, we prepared the protein by using the *Protein Preparation Wizard* workflow implemented in Maestro as well. According to this procedure hydrogen atoms were added and charges and protonation states were assigned titrating the protein at pH 7. Steric clashes were relieved by performing a small number of minimization steps, not intended to minimize the system completely. In our study, the minimization (OPLS force field) was stopped when the RMSD of the non-hydrogen atoms reached 0.30 Å.

Ligands preparation. We employed a proprietary chemical collection available in the D3 Department at IIT for the virtual screening. This collection contains a diverse and nonredundant set of ~15,000 molecules, which are readily available for experimental testing. We prepared the molecules using *LigPrep* software implemented in Maestro. Firstly, we added hydrogens and generated ionization states at pH 7.4 ± 0.5 . Then we generated various tautomers and all stereochemical isomers when not specified in the input structures. For each structure containing a ring moiety, the low-energy conformation was computed and retained. Lastly, a short

minimization step was carried out in order to relax the 3D structure of each molecule.

At this point we filtered the resulting database in order to discard molecules that are not endowed of drug-like properties. For example, molecules poorly soluble or too large in size were excluded from the final database. To do so, we firstly computed the ADME descriptors for each molecule using *QikProp* software implemented in Maestro. As filter, we discarded all the molecules that do not respect the Lipinsky's rule of five. In particular, this rule states that molecules having no more than 5 H-bond donors, no more than 10 H-bond acceptors, a MW less than 500 Da and a partition coefficient (logP) not greater than 5 have properties that make them likely to be a candidate drug.

Virtual screening. The first step of the virtual screening consisted on the identification of the pocket on the RhoJ surface to center the grid. To do so, we firstly analyzed the crystallographic complex of Cdc42 and the CRIB domain of PAK6 (PDB code 2ODB). In fact, due to high sequence similarity of both Cdc42-RhoJ and the CRIB domain of PAK6-PAK1, it is reasonable to think that the interactions between Cdc42 and PAK6 are also conserved in RhoJ and PAK1. In Cdc42-PAK6 complex we identified a pocket on the Cdc42 surface in proximity of Ser71, Arg68, Tyr64 (see Figure 6.1). Interestingly, the Trp40 of PAK6 binds this pocket interacting via an H-bond with Ser71, suggesting that this interaction might be one of the anchor points that mediate the protein-protein interaction. Since this surface cavity is conserved also in RhoJ, displaying the same residues than in Cdc42 and also the tryptophan is conserved in PAK1, we used this pocket to center the grid. The cubic grid box was centered on Ser89 of RhoJ having a side of the box of 26Å. We used *Glide* to perform the virtual screening, using Single Precision and retaining one pose for each ligand.

1.3. Preliminary results

After a visual inspection of the binding mode of the best scored molecules, we selected 50 compounds to be sent for biological evaluation. Among these, 8 molecules resulted active with an IC₅₀ of low-mid micromolar range. Interestingly, the binding mode of these molecules shows a common feature that is the interaction

via an H-bond with the Ser89. Also, these molecules bear a planar ring, single or condensed that overlap nicely with the Trp40 of PAK6 in 2ODB (Figure 6.1). Taken together these results confirm that the interaction with Ser89 is an important anchor point to allow the binding of protein-protein inhibitors.

Currently in our labs, we are optimizing our compounds by small chemical modifications in order to increase the affinity with the receptor and to reach an optimal pharmacokinetic profile to study them in vivo.

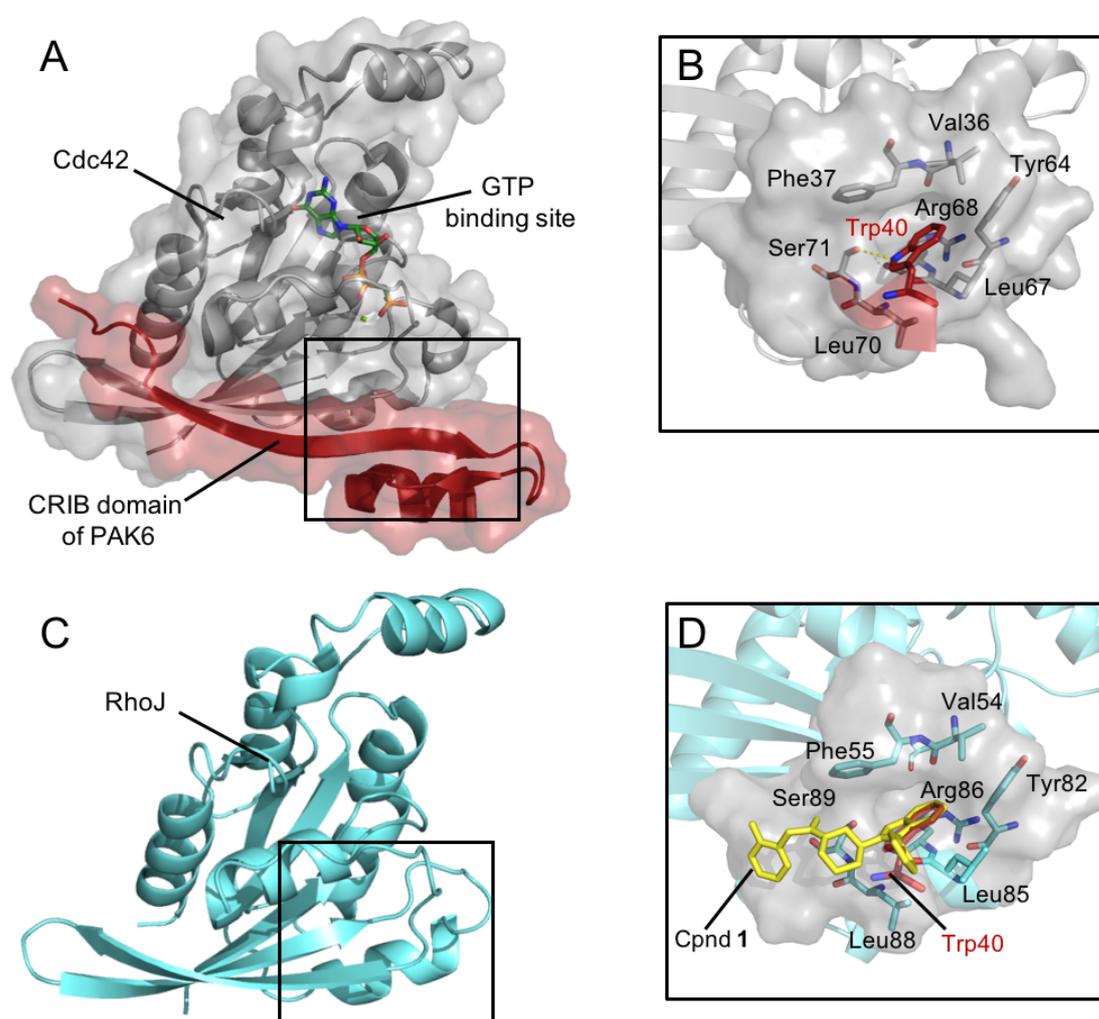


Figure 6.1. A) *Cdc42* protein (grey) in complex with the CRIB domain of PAK6 (red) and a GTP-like molecule (green) bound the GTP binding site. PDB code 2ODB. B) Close up view of the surface binding pocket of *Cdc42*. The Trp40, belonging to PAK6 interacts via an H-bond with the Ser71 of *Cdc42*. C) Homology model of *RhoJ*. D) Binding mode of our hit compound (yellow). The scaffold of our compound overlaps well with the Trp40 of PAK6.

2. Acid Ceramidase

2.1. Introduction

Acid ceramidase (AC) is a cysteine amidase that hydrolyses the proapoptotic lipid-derived messenger ceramide, a central molecule in sphingolipid metabolism and signaling. The expression of AC is abnormally high in several tumors, suggesting that dysregulation of ceramide metabolism could be linked to cancer development and progression.^{240, 241} Indeed, AC overexpression has been detected in various types of human cancer (e.g., melanoma, head and neck, prostate and colon cancer). Moreover, autoantibodies against AC have been found in serum of melanoma patients, suggesting that AC might also be utilized as biomarker for disease prognosis.²⁴² There is also evidence that overexpression of AC may be related with higher resistance to pharmacological induction of apoptosis,^{243, 244} while AC inhibition sensitizes cancer cells to the action of anti-cancer therapeutics.²⁴³ Taken together, there is a growing body of evidence that indicates AC inhibition as an effective and innovative therapeutic approach to treat melanoma, as well as other cancers.^{245, 246}

Recently, potent inhibitors of AC have been identified,^{247, 248} such as the benzoxazolone carboxamide compounds. These inhibitors are quite potent, being active in the nanomolar range. Notably, this class of inhibitors block AC through the formation of a covalent adduct with the catalytic Cys-143. This covalent ligand-target interaction is due to the presence of an urea group in the benzoxazolone carboxamide derivatives. The urea is cleaved to generate the covalent AC-inhibitor complex.²⁴⁹ Although promising, these benzoxazolone carboxamides derivatives however suffer from limited stability, which may hinder their future development towards clinically viable therapeutic agents. In order to understand what chemical features are necessary for potent AC inhibition, while improving chemical stability, we developed a set of different scaffolds. In analogy to their template benzoxazolone carboxamide derivatives, such compounds covalently inhibit AC. The general pharmacophore of this compounds combines 6+5 fused rings heterocycle with different aliphatic chains, linked via an urea functional group. In this context, I applied molecular modeling techniques focusing on the urea moiety, aiming at

understanding whether a particular urea spatial orientation is preferred for AC inhibition.

2.2. Computational methods

We analyzed six compounds having a phenylbutyl moiety and benzimidazole derivative linked via an urea functional group (compound **8**, **10**, **19**, **20**, **21** and **22**). We performed a force field-based coordinate scan of dihedral angle ψ formed by C2-N3-C3'-O3' atoms. To do this we employed the *coordinate scan* tool implemented in Maestro software, version 10.4. First the scaffold of the six compounds were manually designed replacing the phenylbutyl moiety with a methyl group. Then the 360° coordinate scan was performed using the OPLS 2005 force field in water and incrementing the dihedral angle of 10°.

2.3. Results

On the most potent and promising AC inhibitors (compounds **8**, **10**, **19**, **20**, **21** and **22**), we performed a coordinate scan analysis as a function of the dihedral angle ψ centered along the N3-C3' axis of the C2-N3-C3'-O3' atoms of the urea moiety. In this way, we could identify the most favorable conformations assumed by the urea, which undergoes nucleophilic attack by the Cys-143 residue in AC, generating the covalent enzyme-ligand adduct. In all compounds, we detected two energetic minima for ψ . One located at $\psi = 180^\circ$, where the oxygen atom of the urea group at position 1 is oriented toward the 'outer' side of the molecule (this conformation is named as *eso*, Figure 6.2) One located at $\psi = 360^\circ$, where the oxygen atom points toward the 'inner' side of the molecule (this conformation is named as *endo*, Figure 6.2). Interestingly, the most active compounds **8**, **10** and **22** (IC₅₀ of 0.9 ± 0.2 , 2.5 ± 0.1 and 1.0 ± 0.4 nM, respectively) prefer the *eso* conformation, while the others prefer *endo* conformations. Compound **8** is highly stabilized in *eso* thanks to the formation of an intramolecular H-bond between the amino group in position 2 and the oxygen atom of the urea. However, *eso* conformations are also preferred in **10** and **22**, in the absence of this H-bond, indicating that such conformation is stabilized by the nature of the central core. In fact, the less active compounds **19**, **20** and **21**, (IC₅₀ of >10000,

21.0 ± 6.9 and 13.9 ± 3.9 nM, respectively) prefer the *endo* conformation in which all of them form an intramolecular H-bond (Figure 6.2). These results thus suggest that the nucleophilic attack performed by Cys-143 is favored when the urea assumes the *eso* conformations, which therefore return more potent compounds, while amide bond distortion might be relevant for amide bond hydrolysis. The minimal energy conformation assumed by ψ and, in particular, the stabilization of the *eso* conformation seems to be an important prerequisite for the rational design of a potent compound against AC.

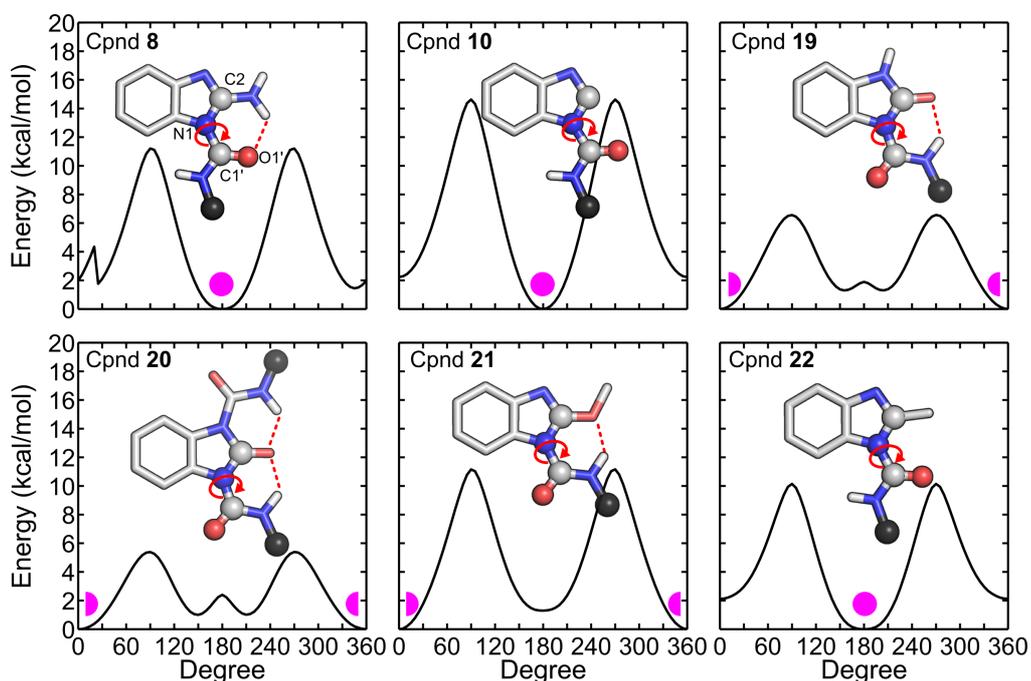


Figure 6.2. Coordinate scan plot for compounds 8, 10, 19, 20, 21 and 22. The pink dot in the plot locates the absolute minimum conformation of ψ . It can be in *eso* when ψ is 180° or in *endo* when ψ is 0° or 360° . The 2D structure depicted in each graph represents the minimum conformation.

Appendix A

Supporting Material

HRD motif as the central hub of the signaling network for activation loop autophosphorylation in Abl kinase

Supplementary text

Structural model

As reported in the paper, we employed 1OPL chain B X-ray structure (resolution 3.4 Å) to model the SH2-KD_{out} and KD_{out} systems. This structure has been also employed in recent computational studies that demonstrated the importance of the SH2 domain in favoring the catalytic activity of Abl kinase,^{129, 130} Notably, the KD of 1OPL chain B matches well the 2HZI structure (another crystal of Abl, with better resolution), returning an RMSD of 0.4 Å. The backbone and the side chains of residues located in the most important structural motifs of KD overlap very well (namely, the A-loop, the DFG motif, the HRD motif and the αF helix), demonstrating only subtle differences in the exposed side chains (Fig. S13). Notably, we added the missing residues from the X-ray 1OPL chain B, namely the SH2-KD linker (residues 238-251 Abl isoform 1b numbering) and the terminal part of the αI helix (residues 519-531) to simulate the entire structure of the SH2-KD complex employed in the experimental studies performed by Lamontanara et al.¹²⁸

The SH2-KD linker is an important component of kinases. When Abl is in an auto-inhibited form (as in the PDBid 1OPK), the SH2-KD linker interacts with domain residues stabilizing the closed form of the protein. Conversely, in the top-hat conformation, the SH2-KD linker is highly flexible and exposed to the solvent, as evidenced by SAXS data. Despite the linker does not directly interact with the kinase, the presence of the linker within the SH2-KD_{out} system is essential to simulate a realistic model of the SH2-KD complex. Indeed, the linker could act as a leash, modulating the degree of freedom of both the SH2 domain and the KD, thus acting on the motions of the complex. The absence of this important structural feature could have led to unnatural or unrealistic motions that we wanted to avoid. The missing residues of the linker were added manually using as reference the FASTA sequence. Then, such portion was submitted to conformational search using the *refinement* tool of Prime where only the lowest energy conformer was retained (Fig. S11). The terminal portion of the αI helix was built by homology modeling, using the PDB code 2F4J as a template. Also, the Prime software was employed for this task. To reproduce the wild-type structure of SH2-KD_{out} and KD_{out}, the N382 was replaced with an aspartate, while for the KD_{in} system the P415 was replaced

with one histidine. The final structure of the SH2-KD_{out} system includes residues from 138 to 531, while for the other systems (i.e. KD_{out}, KD_{in}, D440A-KD_{out}, Closed-KD_{out}) the final structures include residues from 248 to 531.

Hydrogen atoms were added to each system. Subsequently, each system was immersed in a water box where the distance between the solute and the edge of the box was set to 10 Å. To neutralize the overall charge of the systems, a number of water molecules equal to the protein net charge were replaced by Na⁺ counter-ions. For all these three procedures, we used the tLEAP program of the AMBER 10 package.

Classical molecular dynamics simulations

To run classical MD simulations, we adopted a two-step protocol that includes: i) equilibration phase; ii) production phase. For the equilibration phase, each system was minimized using a steepest-descent minimization algorithm. Then, each system was heated from 0 to 300 K in 100 ps in NVT ensemble, keeping the C α atoms fixed in their original positions, using an harmonic force constant of 1000 kJ mol⁻¹ nm⁻². Subsequently, 5 ns NPT simulation was carried out for each heated system at 300 K and 1 bar, maintaining C α constrained. At this point, all the constraints were removed and the production phase of MD was carried out in the NPT ensemble. However, we considered the first 100 ns of the production phase as part of the initial equilibration and for this reason it was discarded from subsequent analyses. Bonds involving hydrogen atoms were restrained to their equilibrium length with the P-LINCS algorithm.

All simulations were performed using a short-range neighbor list cutoff of 9 Å, whereas long-range electrostatics were treated with the Particle Mesh Ewald method using a Fourier grid spacing of 1.6 Å and a cubic spline interpolation. This set up was employed also in a number of published works presenting microsecond-long MD simulations. Periodic boundary conditions in the three directions of Cartesian space were applied and a time step of 2 fs was set for integrating the equations of motion. All simulations were performed with Gromacs 4.6 software. A Parrinello-Rahman barostat and a velocity-rescaling thermostat were employed with a relaxation time τ of 2 ps and 0.1 ps, respectively.

All analyses were performed on snapshots saved every 50 ps from the original trajectory for a total of ~36000 (SH2-KD_{out}), ~52000 (KD_{out}), ~22000 (Replica1-KD_{out}), ~22000 (Replica2-KD_{out}), ~26000 (KD_{in}), ~26000 (D440A-KD_{out}), ~26000 (closed-KD_{out}) frames.

Analysis of structural data from MD

To evaluate the stability of the systems during our MD simulations, we computed the RMSD of the protein backbone employing `g_rms` tool of Gromacs 4.6 package. In addition, for the closed-KD_{out} system we performed a deeper analysis on the myristate binding pocket to ensure that the structure of the cavity did not undergo major perturbations after the removal of the co-crystallized myristic acid. The RMSD of the C α atoms of the pocket, which is formed by the residues around 8 Å from the myristic acid, returned a value of 1.19 ± 0.33 Å, indicating an overall stability of this cavity. Also, we computed the contact map of the C α atoms of the pocket using `g_mdmat` tool of Gromacs 4.6 package (Fig. S14). This analysis further confirms the overall stability of this cavity with a preservation of the overall structure of the pocket in our simulations, even in the absence of the co-crystallized binder.

The principal component analysis (PCA) is a method employed to describe long trajectories in terms of a small number of variables, called principal components or essential degrees of freedom.¹⁶⁶ Initially, we removed translation and rotation of the whole protein by superimposing each snapshot of the MD simulations onto the starting frame (i.e. X-ray structure). Then, we computed the 3N dimensional covariance matrix, C , of the atomic fluctuations. The elements of C are defined as:

$$c_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle$$

where x_i and x_j are the atomic coordinates and the brackets $\langle \rangle$ indicate the average over time. Finally, a set of eigenvectors and eigenvalues were obtained through the diagonalization of the symmetric matrix C . The eigenvectors represent the directions in a multidimensional space, describing independent modes in atomic motions, while the eigenvalues represent their corresponding amplitudes. As a result, the first few PC modes characterize the dominant motions of the protein in each trajectory. The

most flexible parts of KD (i.e. the first 13 and the last 15 residues of the N- and C-terminal tail, respectively) were discarded from the analysis in order to avoid the associated noise. The PCA was performed by using `g_covar` and `g_anaeig` tools of Gromacs 4.6 package.

Classical metadynamics simulations

In the present work, the well-tempered metadynamic simulations was coupled with multiple walkers technique in order to evaluate the energetic cost due to the A-loop partial closure observed during the unbiased MD simulation.¹⁶⁸

In detail, we run two well-tempered metadynamic simulations (i.e. two walkers) for each system, employing two different A-loop starting configurations: 1) the phosphorylatable fully-open conformation, in which CV (i.e. the distance between the phenolic oxygen of Y412 and the COM of carboxylate atoms of D₄₄₀) is ~ 12 Å; 2) the non-phosphorylatable partially closed conformation, in which CV is ~ 3 Å. For the first walker we extracted a frame from the respective equilibrated unbiased MD simulations. For the second walker, instead, we employed two different procedures for extracting the starting frame. For KD_{out} system we extracted a frame from the unbiased MD simulations. Instead, for SH2-KD_{out} and KD_{in} systems we first performed a classical metadynamics simulation to enable the closure of the A-loop and to obtain a conformation in which CV is ~ 3 Å. In the classical metadynamics simulation the bias added to a defined collective variable S (i.e. CV) can be written as

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

where σ_i is the width of the Gaussian, while ω is the ratio between the Gaussian height (W) and the deposition stride (τ_G). In our simulations we set $\sigma_i = 0.3$ Å, $W = 2$ kJ/mol and $\tau_G = 500$.

Cut-off definition for Y412 conformations

In the present work, we adopted a specific metric for discriminating the different Y412 conformations observed during the MD simulations. Such metric consists of two parameters: i) the RMSD of Y412 in respect to the 1OPL chain B X-ray position ($\text{RMSD}_{\text{Y412}}$); ii) the distance between Y412 and the reference residue D440 along the αF helix, chosen for its high stability along the MD simulations ($d[\text{Y}_{412}\text{-D}_{440}]$). The measurement of the RMSD among several conformations retrieved from the MD trajectories quantifies their similarity in respect to a reference one. In this case a low $\text{RMSD}_{\text{Y412}}$ value stands for Y412 conformations that are similar to the phosphorylatable state adopted in the X-ray 1OPL chain B. However, the drawback in using only the $\text{RMSD}_{\text{Y412}}$ value is that we could not distinguish if an increase of the RMSD is due to a subtle change in the orientation of Y412 side chain or to a displacement of the entire Y412 residue. For this reason we adopted an additional parameter ($d[\text{Y}_{412}\text{-D}_{440}]$) that gives us a clear view on the Y412 localization in respect to the native position. By plotting together the $d[\text{Y}_{412}\text{-D}_{440}]$ and $\text{RMSD}_{\text{Y412}}$ values we obtained a distribution of points that were subsequently submitted to a 2D histogram analysis to get the weighted population distribution. As a result we obtained four main states for KD_{out} and one for $\text{SH2-KD}_{\text{out}}$ and KD_{in} (see Fig. 3.4 in the main text). To separate the phosphorylatable from the non-phosphorylatable states we employed a cut-off on both $\text{RMSD}_{\text{Y412}}$ and $d[\text{Y}_{412}\text{-D}_{440}]$. For choosing the cut-off value we applied the following procedure: first, the $\text{RMSD}_{\text{Y412}}$ and $d[\text{Y}_{412}\text{-D}_{440}]$ values of the three systems (i.e. $\text{SH2-KD}_{\text{out}}$, KD_{out} and KD_{in}) were distributed separately into histograms using a bin width of 0.2 Å. Then, each histogram was fitted through a Gaussian function, (red, blue and green curves for $\text{SH2-KD}_{\text{out}}$, KD_{out} and KD_{in} respectively in Fig. 3.4 of the main body of the paper). Here, we identified the cluster that contains the Y412 conformations that resemble the crystallographic one (i.e. cluster 1/ *Conf-1* in Fig. 3.4). Finally, to quantitatively separate such cluster from the other ones, we used as $\text{RMSD}_{\text{Y412}}$ cut-off the middle value between the tails of first two blue gaussians, (starting from the bottom of Figure 3.4) that is 3.5 Å. Likewise, the $d[\text{Y}_{412}\text{-D}_{440}]$ cut-off is the middle value between the tails of second and the third blue gaussians (starting from the left of Fig. 3.4) that is 8 Å.

Supplementary figures

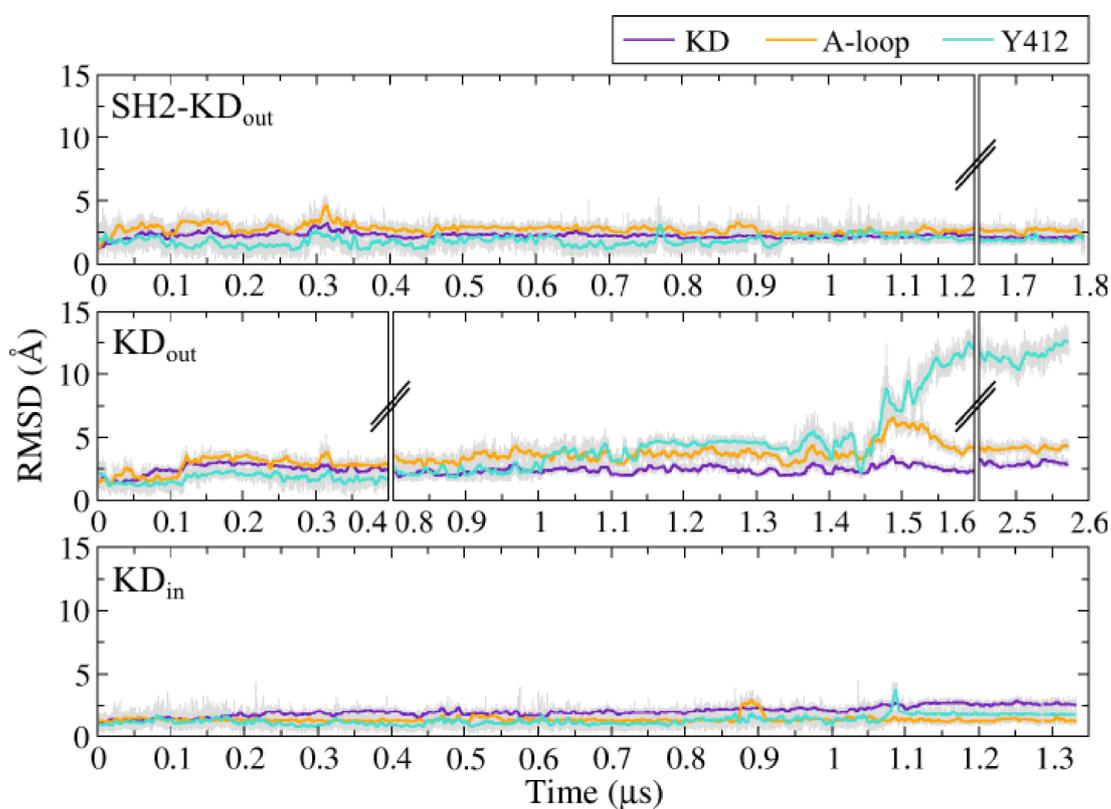


Figure S1. Time evolution of the RMSD for the backbone atoms of KD (violet), A-loop (orange), and Y412 (turquoise) of SH2-KD_{out} (upper panel), KD_{out} (central panel), and KD_{in} (lower panel). The average RMSD values for the KD are 2.26 ± 0.20 Å (SH2-KD_{out}), 2.58 ± 0.33 Å (KD_{out}) and 2.07 ± 0.36 Å (KD_{in}), for the A-loop are 2.77 ± 0.35 Å (SH2-KD_{out}), 3.71 ± 0.79 Å (KD_{out}) and 1.37 ± 0.31 Å (KD_{in}), while for the Y412 are 1.89 ± 0.51 Å (SH2-KD_{out}), 6.11 ± 4.37 Å (KD_{out}) and 1.31 ± 0.52 Å (KD_{in}).

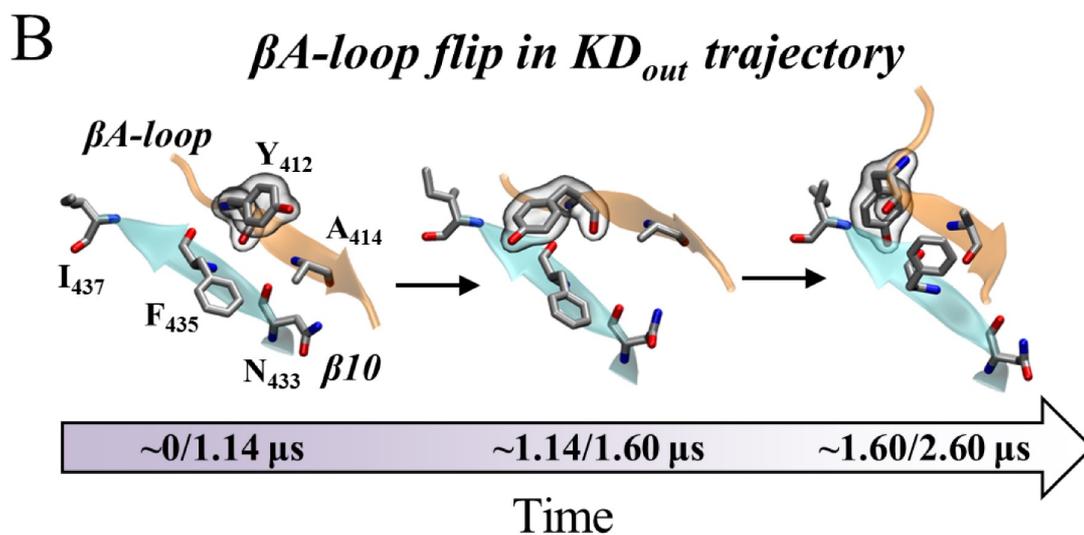
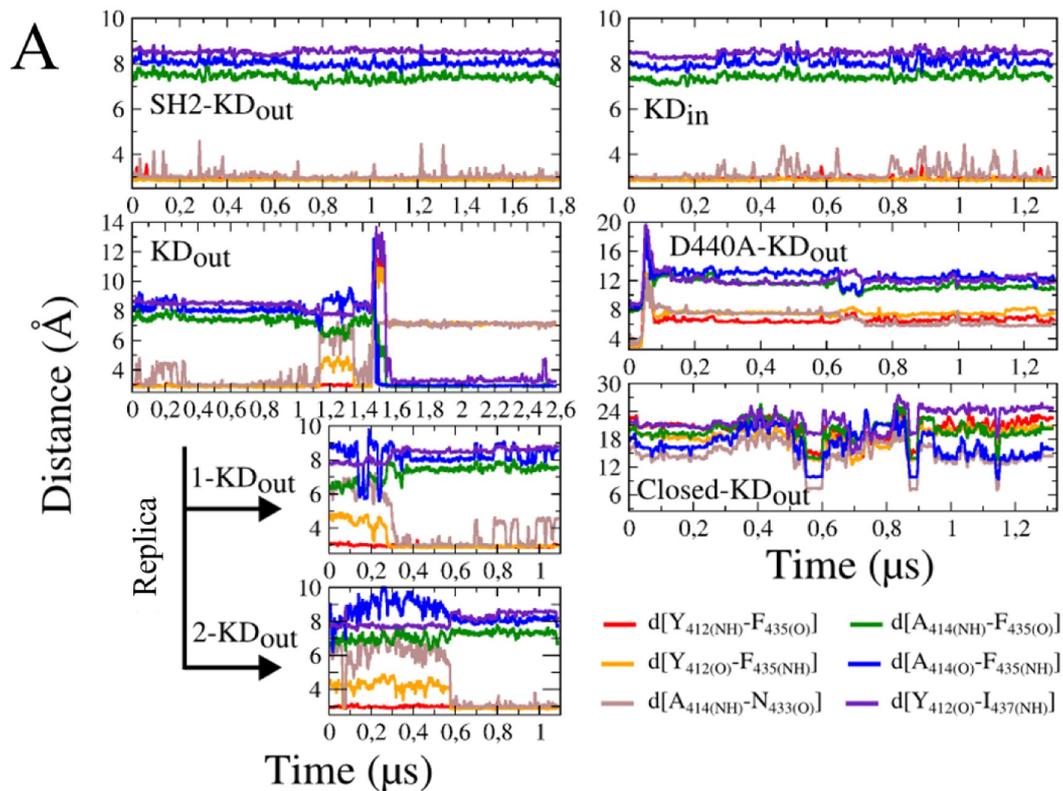


Figure S2. A) Representation of time evolution distances of $d[\text{Y}_{412(\text{NH})}-\text{F}_{435(\text{O})}]$ (red), $d[\text{Y}_{412(\text{O})}-\text{F}_{435(\text{NH})}]$ (orange), $d[\text{A}_{414(\text{NH})}-\text{N}_{433(\text{O})}]$ (brown), $d[\text{A}_{414(\text{NH})}-\text{F}_{435(\text{O})}]$ (green), $d[\text{A}_{414(\text{O})}-\text{F}_{435(\text{NH})}]$ (blue), $d[\text{Y}_{412(\text{O})}-\text{I}_{437(\text{NH})}]$ (violet) of the 7 simulated systems. The crystallographic H-bond network between β 10 and β A-loop is formed by $d[\text{Y}_{412(\text{NH})}-\text{F}_{435(\text{O})}]$, $d[\text{Y}_{412(\text{O})}-\text{F}_{435(\text{NH})}]$ and $d[\text{A}_{414(\text{NH})}-\text{N}_{433(\text{O})}]$. The flip of β A-loop observed only in KD_{out} trajectory leads to the formation of a new H-bond network between β 10 and β A-loop formed this time by $d[\text{A}_{414(\text{NH})}-\text{F}_{435(\text{O})}]$, $d[\text{A}_{414(\text{O})}-\text{F}_{435(\text{NH})}]$ and $d[\text{Y}_{412(\text{O})}-\text{I}_{437(\text{NH})}]$. In both D440A-KD_{out} and closed-KD_{out} the H-bond network between β 10 and β A-loop is not formed. B) Schematic representation of flip observed during the KD_{out} trajectory. The phosphorylatable Y412 is represented in surface.

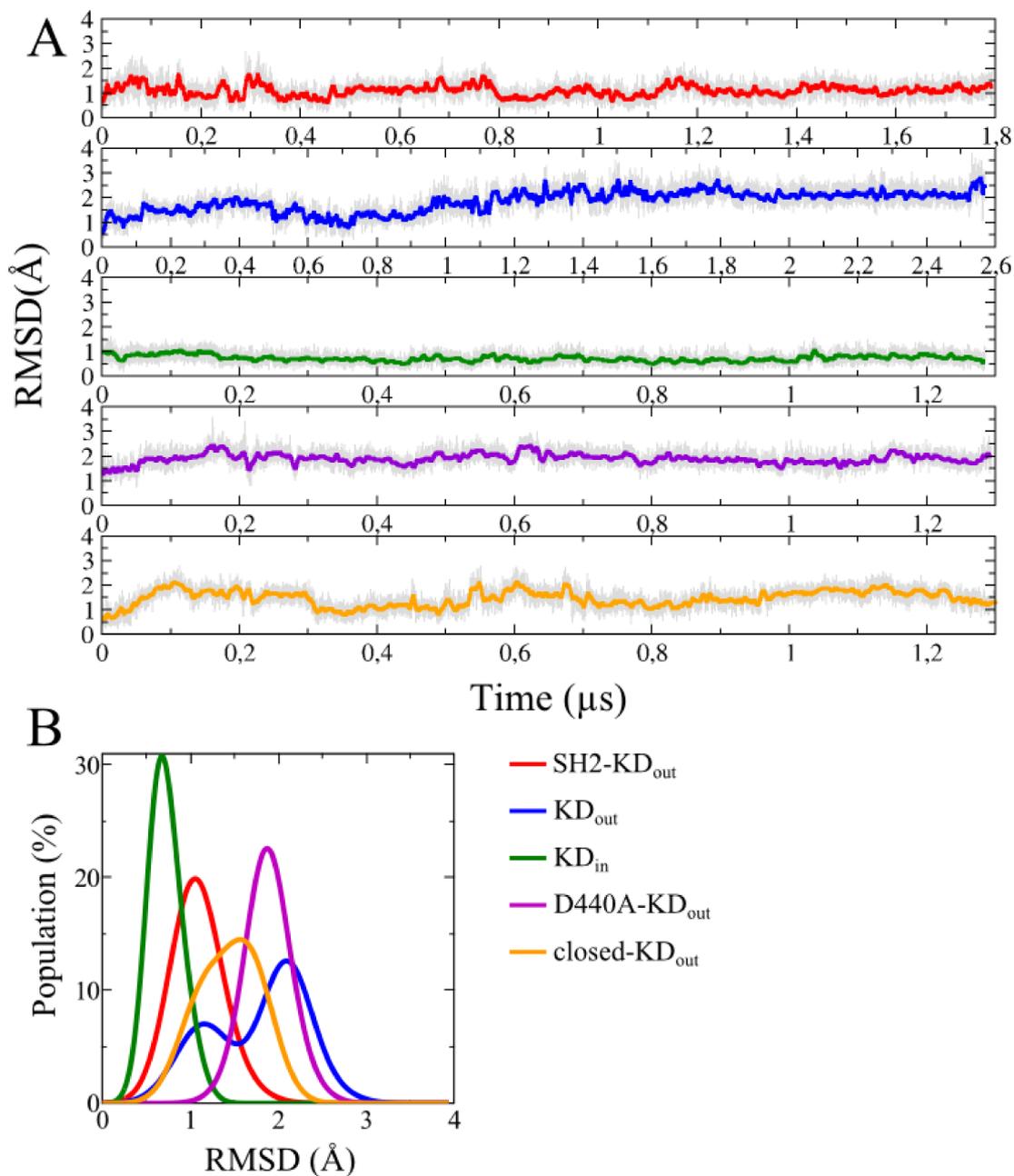


Figure S3. A) Time evolution of RMSD backbone atoms of the HRD motif for SH2-KD_{out} (red), KD_{out} (blue), KD_{in} (green), D440A-KD_{out} (purple) and closed-KD_{out} (orange) trajectories. The average RMSD values for the HRD motif are 1.08 ± 0.29 Å (SH2-KD_{out}), 1.80 ± 0.54 Å (KD_{out}), 0.70 ± 0.10 Å (KD_{in}), 1.90 ± 0.26 Å (D440A-KD_{out}) and 1.46 ± 0.36 Å (Closed-KD_{out}). B) Normalized frequency distributions of the RMSD values.

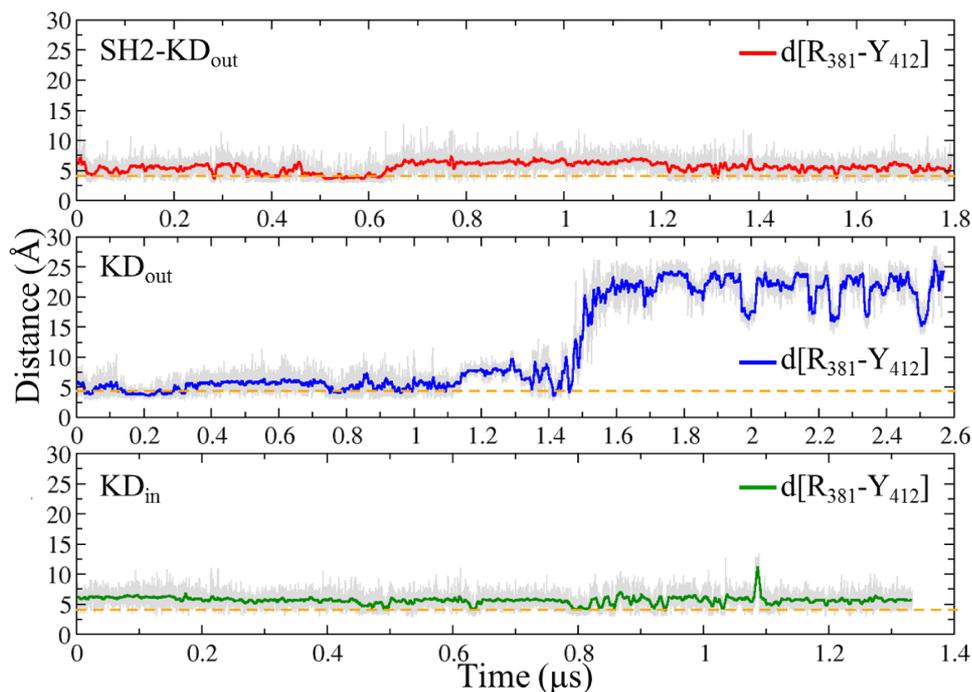


Figure S4. Time evolution of $d[R_{381}-Y_{412}]$ respectively for SH2-KD_{out} (upper panel), KD_{out} (middle panel) and KD_{in} (lower panel). The dashed orange line represents the crystallographic value of 1OPL chain B structure (3.90 Å). The average value of $d[R_{381}-Y_{412}]$ are 5.61 ± 1.21 Å (SH2-KD_{out}), 12.37 ± 8.02 Å (KD_{out}) and 5.69 ± 1.08 Å (KD_{in}).

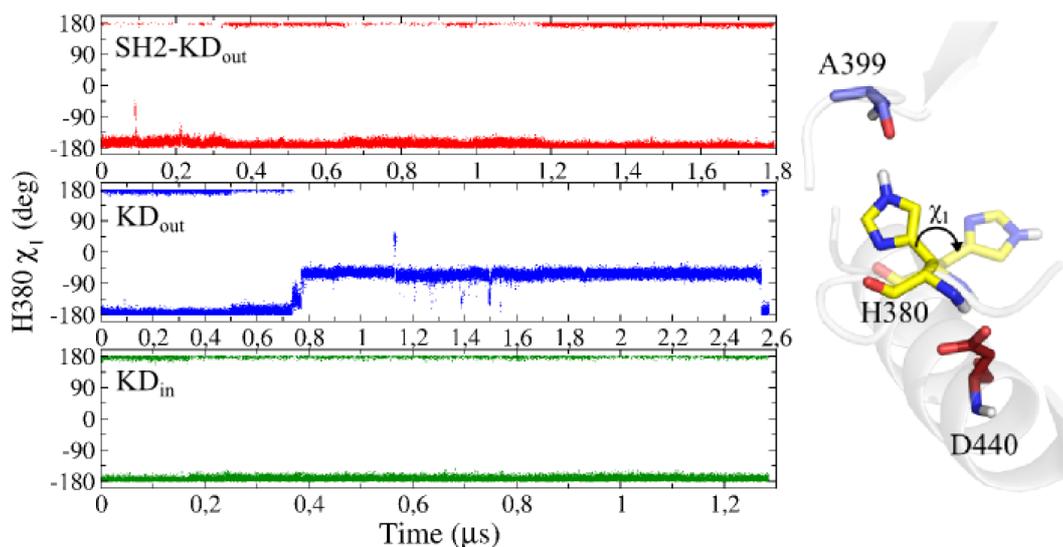


Figure S5. Values along time of χ_1 dihedral angle of H380 of SH2-KD_{out} (upper panel), KD_{out} (middle panel) and KD_{in} (lower panel).

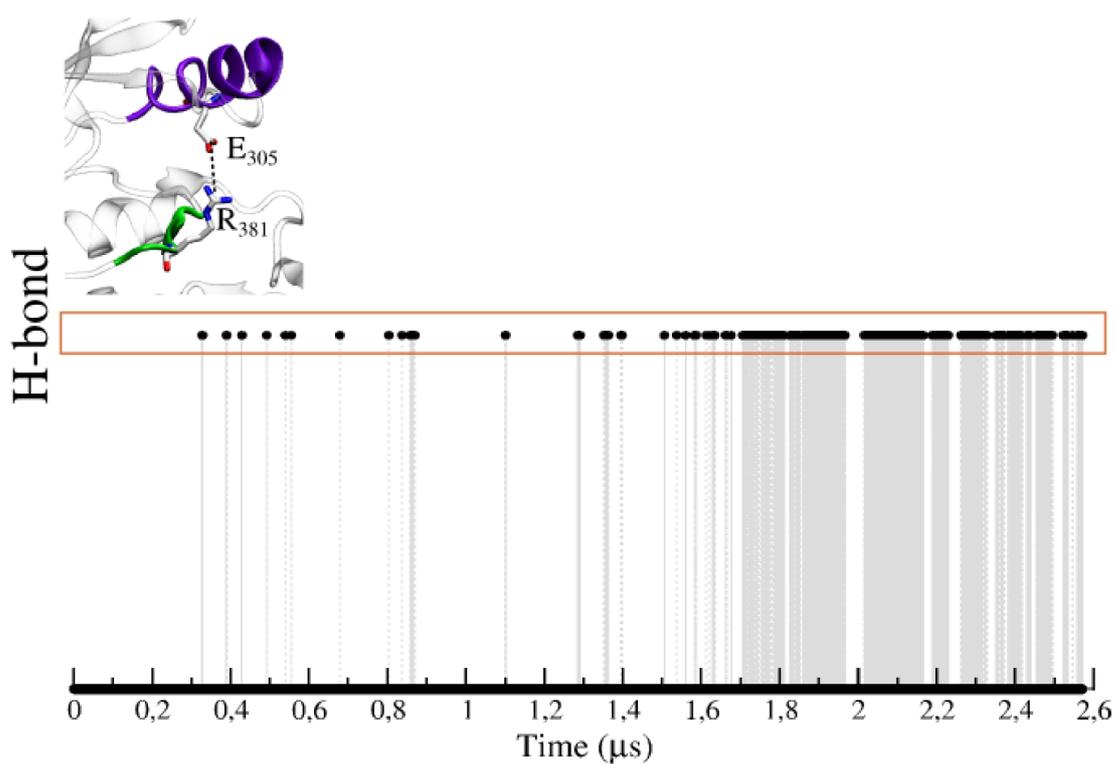


Figure S6. Representation of the frequency of the H-bond between the E305 along the α C helix and the R381 located at the HRD motif in the KD_{out} trajectory. The black dots, evidenced by the orange square, indicate the presence of an H-bond between such residue pair that becomes more frequent in the last part of the KD_{out} trajectory (i.e. when the Y412 is in Conf-4).

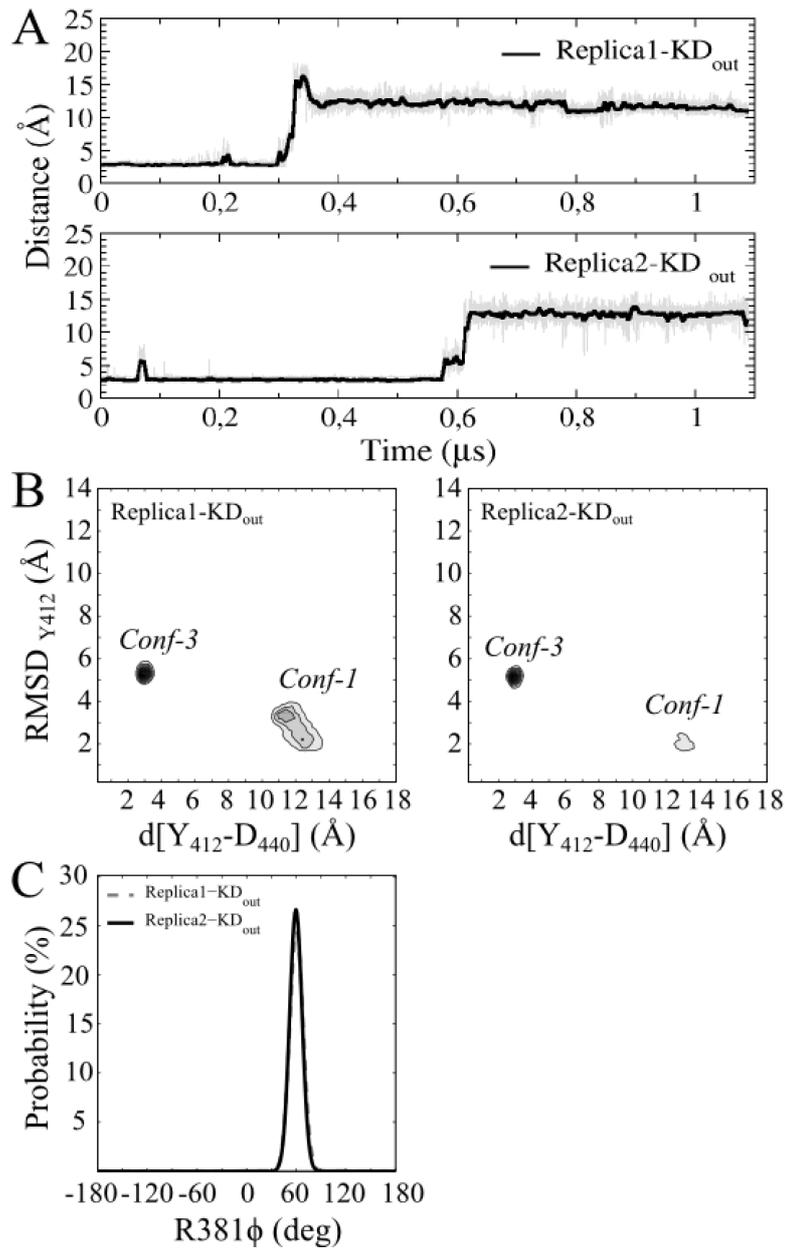


Figure S7. *A*) Time evolution of $d[Y_{412}-D_{440}]$ in Replica1-KD_{out} (upper panel) and Replica2-KD_{out} (lower panel). Both trajectories start in Conf-3, where the $d[Y_{412}-D_{440}]$ is ~ 3 Å, but at ~ 0.3 and 0.55 μs the distance elongates reaching ~ 12 Å for Replica1-KD_{out} and Replica2-KD_{out}, respectively. *B*) 2D plot of Replica1-KD_{out} (left panel) and Replica2-KD_{out} (right panel) trajectories data points projected along $d[Y_{412}-D_{440}]$ –X axes– and the RMSD_{Y412} –Y axes–. In both systems, the Y412 visited 2 states: the starting Conf-3 and the phosphorylatable Conf-1. *C*) Normalized frequency distribution of R381 ϕ dihedral angle in Replica1-KD_{out} (grey dashed line), Replica2-KD_{out} (black line) trajectories.

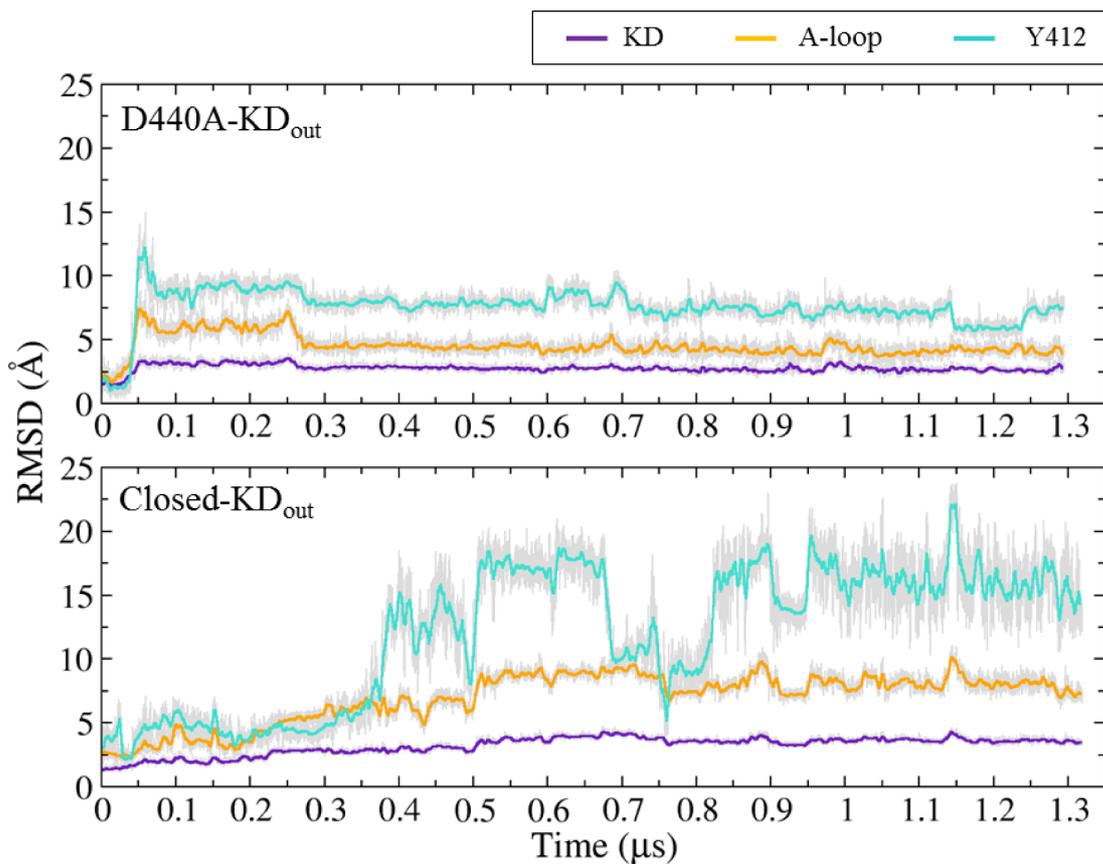


Figure S8. Time evolution of the RMSD of the backbone atoms of KD (violet), A-loop (orange) and Y412 (turquoise) of D440A-KD_{out} (upper panel) and Closed-KD_{out} (lower panel) simulations. The average RMSD values for the KD are $2.82 \pm 0.26 \text{ \AA}$ (D440A-KD_{out}), $3.25 \pm 0.62 \text{ \AA}$ (Closed-KD_{out}), for the A-loop are $4.60 \pm 0.77 \text{ \AA}$ (D440A-KD_{out}), $7.04 \pm 1.82 \text{ \AA}$ (Closed-KD_{out}), while for the Y412 are $7.81 \pm 0.88 \text{ \AA}$ (D440A-KD_{out}), $11.57 \pm 5.27 \text{ \AA}$ (Closed-KD_{out}).

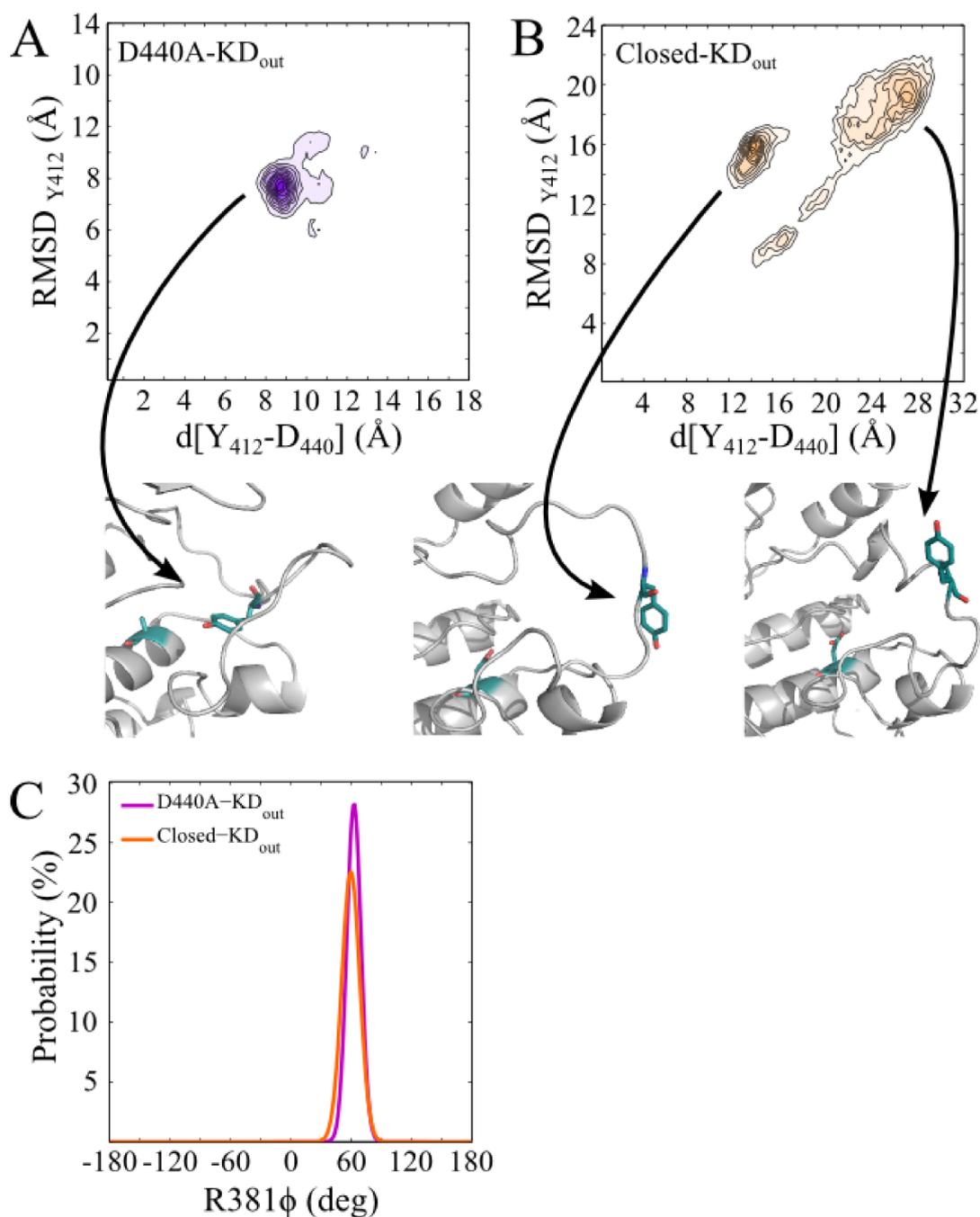


Figure S9. *A)* 2D plot of D440A-KD_{out} MD trajectory projected along $d[Y_{412}-A_{440}]$ – X axes – and $RMSD_{Y_{412}}$ values – Y axes – values. The figure under the plot represents the most populated conformation. *B)* 2D plot of Closed-KD_{out} MD trajectory projected along $d[Y_{412}-D_{440}]$ – X axes – and $RMSD_{Y_{412}}$ values – Y axes – values. The figures under the plot represent the most populated conformations. *C)* Frequency distribution of R381 ϕ dihedral angle in D440A-KD_{out} (violet) and Closed-KD_{out} (orange).

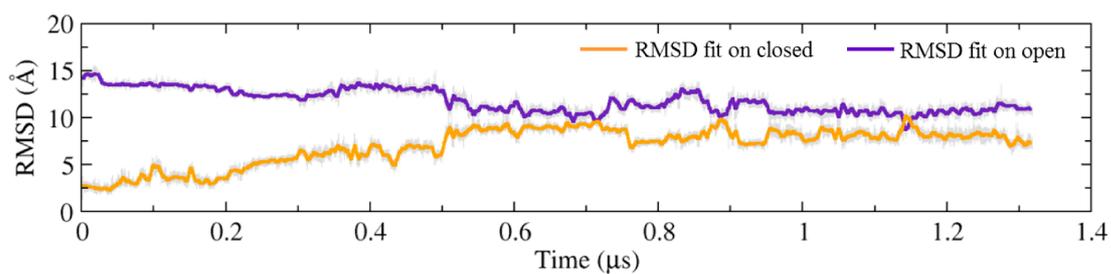


Figure S10. RMSD projection along time of the backbone A-loop respect to the closed conformation, assumed in 1OPJ X-ray (orange) and respect to fully open conformation as in 1OPL X-ray (violet) in Closed-KD_{out} simulations.

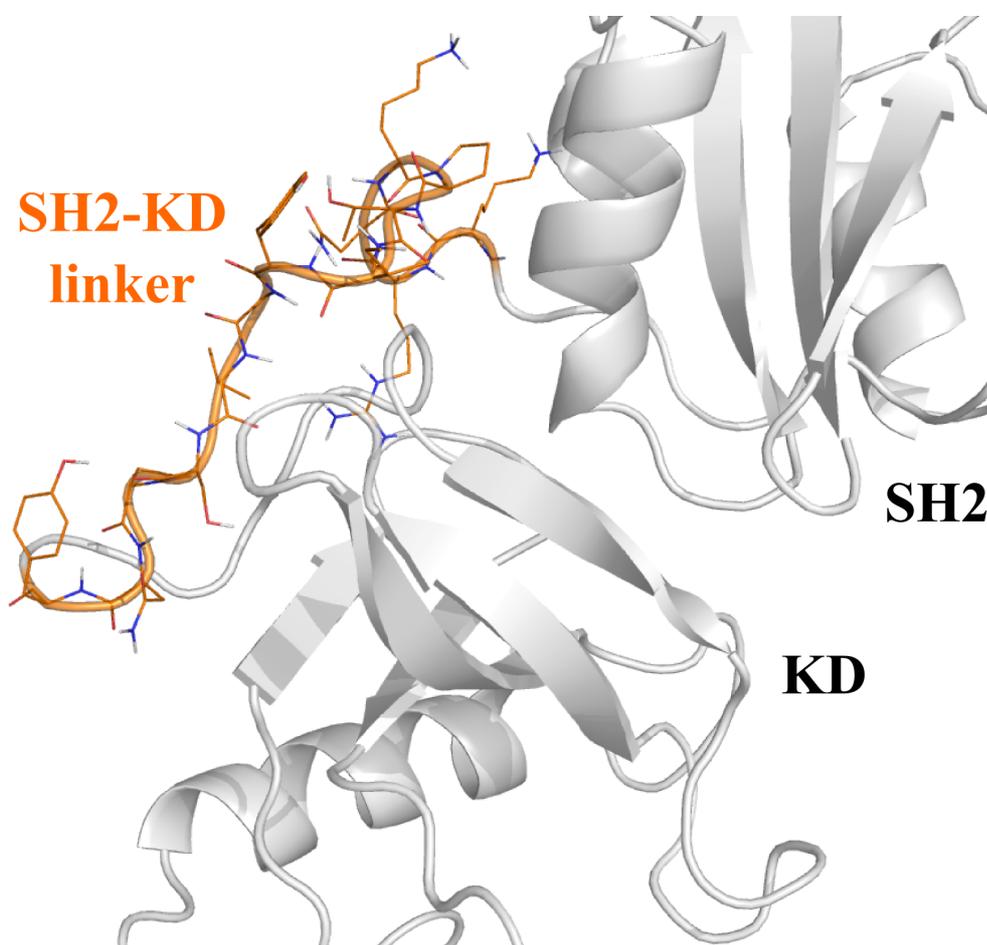


Figure S11. Representation of the modeled SH2-KD linker (orange) of the SH2-KD_{out} system.

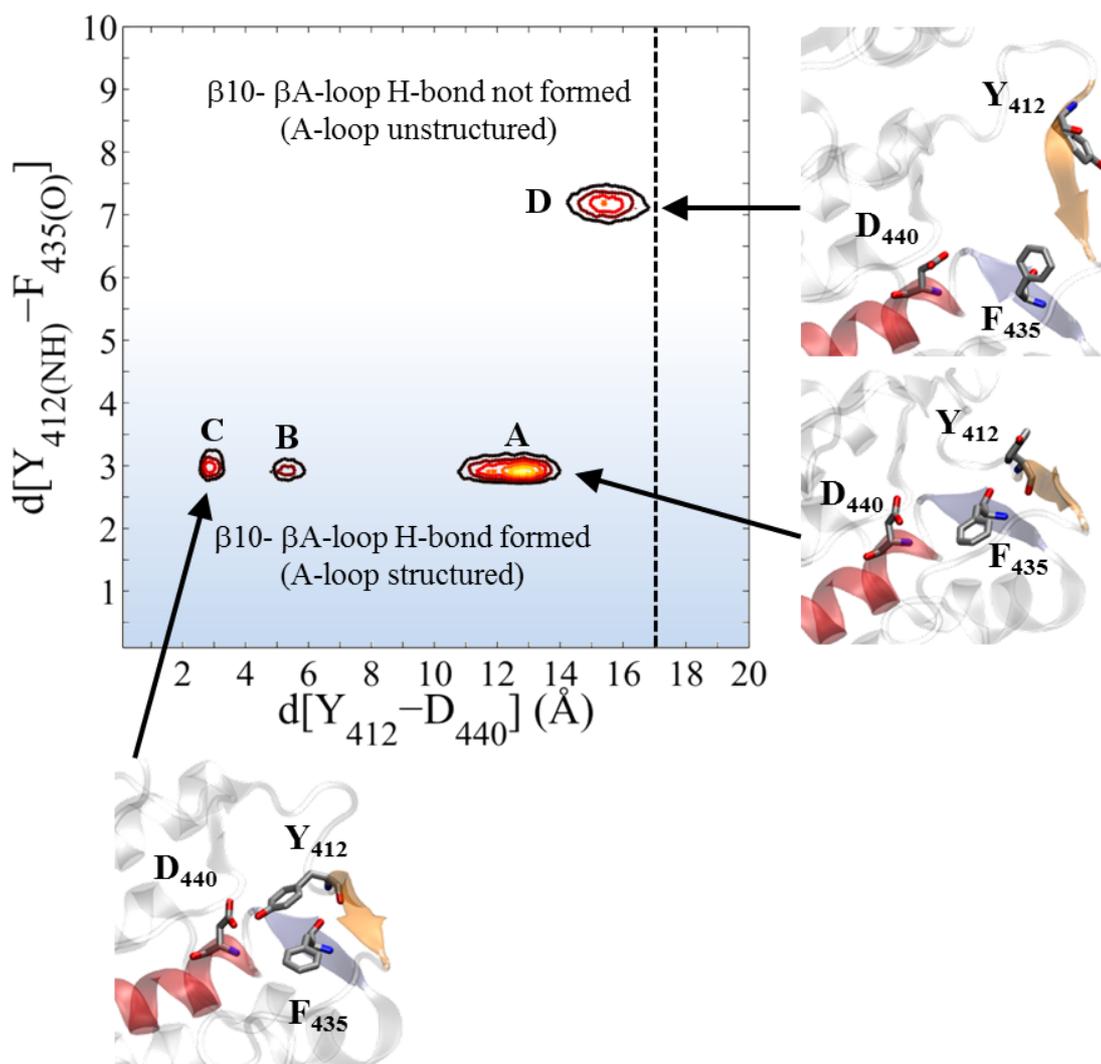


Figure S12. 2D plot of $d[Y_{412}-D_{440}]$ values plotted against $d[Y_{412(NH)}-F_{435(O)}]$ (which is a key H-bond for the formation of the two antiparallel β sheets $\beta 10$ - βA -loop H-bond network) in the KD_{out} trajectory. The A basin corresponds to those conformations in which the A-loop is fully open. The B and C basins correspond to those conformations in which the A-loop is partially closed. The conformations in A, B and C are defined as “structured” (see multiple walkers metadynamics paragraph in methods section) inasmuch the A-loop conserves the overall native assembly. This is evidenced by the proper formation of the $\beta 10$ - βA -loop H-bond network (light blue background). The increase of $d[Y_{412}-D_{440}]$ values is related to the destabilization of the A-loop which assumes “unstructured” conformations (basin D) causing the rupture of the $\beta 10$ - βA -loop H-bond network (white background). Since in this work we focused only in computing the free energy profile of the A-loop fully open \rightarrow partial closed transitions among different Abl systems, we applied a cut-off on $d[Y_{412}-D_{440}]$ (dotted black line) to discard the non-interesting “unstructured” A-loop conformations.

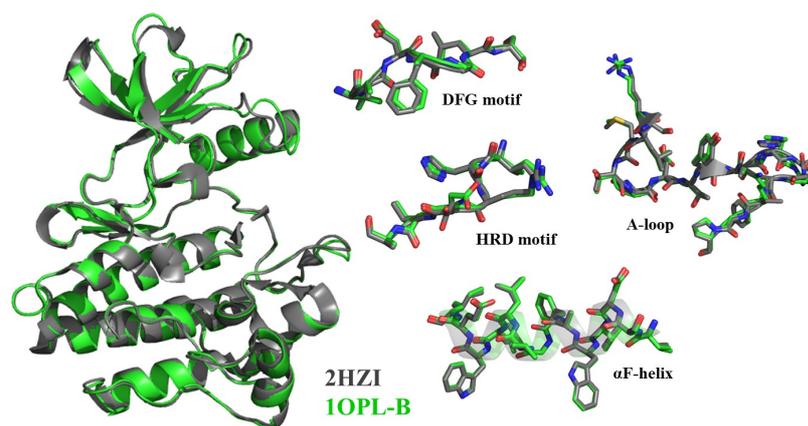


Figure S13. *Overlap between 1OPL chain B X-ray structure (resolution 3.4 Å) and 2HZI X-ray structure (resolution 1.7 Å). On the left side the two kinase domains are overlapped. On the right side, the superposition of key portions of KD (i.e. DFG motif, HRD motif, A-loop and α F-helix) is shown.*

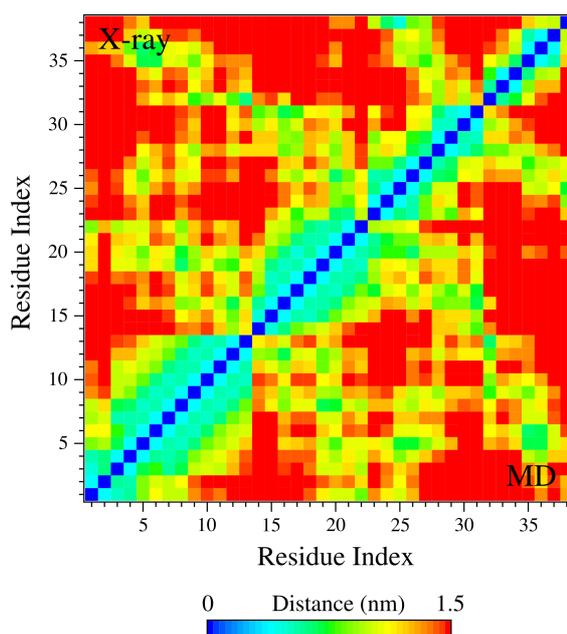


Figure S14. *Contact map of residues of the myristate binding site in 1OPJ X-ray structure (upper triangular matrix) and in the closed- KD_{out} MD simulations (lower triangular matrix). The contact map is calculated for the $C\alpha$ atom around 8 Å from the myristic acid.*

Supplementary table

PDB ID	Chain ID	Ligand ID	Resolution	Mutation	Species	DFG	A-loop	HRD-R(ϕ)	α C-Helix	Domain
1FPU	A	PRC	2.40		M.M.	OUT	CLOSED	(+)77.96°	IN	KD
1FPU	B	PRC	2.40		M.M.	OUT	CLOSED	(+)76.84°	IN	KD
1IEP	A	STI	2.10		M.M.	OUT	CLOSED	(+)87.76°	IN	KD
1IEP	B	STI	2.10		M.M.	OUT	CLOSED	(+)87.89°	IN	KD
1MS2	A	P17	2.60		M.M.	OUT	OPEN	(-)84.41°	IN	KD
1MS2	B	P17	2.60		M.M.	OUT	OPEN	(-)78.82°	IN	KD
1OPJ	A	STI+MYR	1.75		M.M.	OUT	CLOSED	(+)85.47°	IN	KD
1OPJ	B	STI+MYR	1.75		M.M.	OUT	CLOSED	(+)87.48°	IN	KD
1OPK	A	P16+MYR	1.80	D382N	M.M.	OUT	OPEN	(+)73.87°	IN	SH2-SH3-KD
1OPL	A	P16+MYR	3.42	D382N, K29R, E29D	H.S.	OUT	OPEN	(+)73.72°	IN	SH2-SH3-KD
1OPL	B	P16	3.42	D382N, K29R, E29D	H.S.	OUT	OPEN	(+)73.81°	IN	SH2-KD
2E2B	A	406	2.20		H.S.	OUT	CLOSED	(+)81.83°	IN	KD
2E2B	B	406	2.20		H.S.	OUT	CLOSED	(+)77.65°	IN	KD
2F4J	A	VX6	1.91	H396P	H.S.	IN	OPEN	(+)72.68°	IN	KD
2F00	A	P16+MYR	2.27	D382N	H.S.	OUT	OPEN	(+)80.44°	IN	SH2-SH3-KD
2GIT	A	112	1.80		H.S.	IN	Src/like	(+)74.07°	OUT	KD
2GIT	B	112	1.80		H.S.	IN	Src/like	(+)81.24°	OUT	KD
2GIT	C	112	1.80		H.S.	IN	Src/like	(+)80.53°	OUT	KD
2GIT	D	112	1.80		H.S.	IN	Src/like	(+)85.75°	OUT	KD
2G2F	A	112	2.70	H396P	H.S.	Intermediate	OPEN	(-)60.45°	OUT	KD
2G2F	B	AGS	2.70	H396P	H.S.	OUT	OPEN	(-)73.34°	IN	KD
2G2H	A	P16	2.00	H396P	H.S.	Intermediate	OPEN	(-)68.95°	OUT	KD
2G2H	B	P16	2.00	H396P	H.S.	OUT	OPEN	(-)89.37°	IN	KD
2G2I	A	ADP	3.12	H396P	H.S.	IN	OPEN	(-)55.25°	IN	KD
2G2I	B	ADP	3.12	H396P	H.S.	IN	OPEN	(-)66.16°	IN	KD
2GQG	A	1N1	2.40		H.S.	IN	OPEN	(+)69.10°	IN	KD
2GQG	B	1N1	2.40		H.S.	IN	OPEN	(+)65.85°	IN	KD
2HIW	A	7MP	2.20		H.S.	OUT	CLOSED	(+)84.67°	IN	KD
2HIW	B	7MP	2.20		H.S.	OUT	CLOSED	(+)79.19°	IN	KD
2HYY	A	STI	2.40		H.S.	OUT	CLOSED	(+)91.16°	IN	KD
2HYY	B	STI	2.40		H.S.	OUT	CLOSED	(+)81.02°	IN	KD
2HYY	C	STI	2.40		H.S.	OUT	CLOSED	(+)83.65°	IN	KD
2HYY	D	STI	2.40		H.S.	OUT	CLOSED	(+)86.15°	IN	KD
2HZ0	A	GIN	2.10		H.S.	OUT	OPEN/disordered	(+)79.36°	IN	KD
2HZ0	B	GIN	2.10		H.S.	OUT	N.A.	(+)87.26°	IN	KD
2HZ4	A	4ST	2.80		H.S.	IN	OPEN	(+)66.10°	IN	KD
2HZ4	B	4ST	2.80		H.S.	IN	OPEN	(+)66.31°	IN	KD
2HZ4	C	4ST	2.80		H.S.	IN	OPEN	(+)68.76°	IN	KD
2HZI	A	JIN	1.70		H.S.	OUT	OPEN	(+)74.37°	IN	KD
2HZI	B	JIN	1.70		H.S.	OUT	OPEN	(+)76.53°	IN	KD
2HZN	A	KIN	2.70		M.M.	OUT	CLOSED	(+)87.41°	IN	KD
2QOH	A	P3Y	1.95		M.M.	IN	OPEN	(+)78.16°	IN	KD
2QOH	B	P3Y	1.95		M.M.	IN	OPEN	(+)74.38°	IN	KD
2V7A	A	627	2.50	T315I	H.S.	IN	OPEN	(+)83.27°	IN	KD
2V7A	B	627	2.50	T315I	H.S.	IN	OPEN	(+)77.78°	IN	KD
2Z60	A	P3Y	1.95	T315I	M.M.	IN	OPEN	(+)73.73°	IN	KD
3CS9	A	NIL	2.21		H.S.	OUT	CLOSED	(+)85.05°	IN	KD
3CS9	B	NIL	2.21		H.S.	OUT	N.A.	(+)81.36°	IN	KD
3CS9	C	NIL	2.21		H.S.	OUT	N.A.	(+)81.89°	IN	KD
3CS9	D	NIL	2.21		H.S.	OUT	CLOSED	(+)79.28°	IN	KD
3DK3	A	SX7	2.02	Y393F, L445P	M.M.	Intermediate	OPEN	(+)76.79°	IN	KD
3DK3	B	SX7	2.02	Y393F, L445P	M.M.	Intermediate	OPEN	(+)76.90°	IN	KD

3DK6	A	SX7	2.02	Y253F, L445P	M.M.	OUT	OPEN	(+)73.35°	IN	KD
3DK6	B	SX7	2.02	Y253F, L445P	M.M.	OUT	OPEN	(+)76.04°	IN	KD
3DK7	A	SX7	2.01	T315I, L445P	M.M.	Intermediate	OPEN	(+)79.79°	IN	KD
3DK7	B	SX7	2.01	T315I, L445P	M.M.	OUT	OPEN	(+)75.78°	IN	KD
3IK3	A	0LI	1.90	T315I	M.M.	OUT	CLOSED	(+)85.69°	IN	KD
3IK3	B	0LI	1.90	T315I	M.M.	OUT	CLOSED	(+)90.87°	IN	KD
3K5V	A	STI+STJ	1.74		M.M.	OUT	CLOSED	(+)87.70°	IN	KD
3K5V	B	STI+STJ	1.74		M.M.	OUT	CLOSED	(+)86.12°	IN	KD
3KF4	A	B90	1.90		M.M.	IN	OPEN	(+)80.64°	IN	KD
3KF4	B	B90	1.90		M.M.	IN	OPEN	(+)78.87°	IN	KD
3KFA	A	B91	1.22		M.M.	OUT	CLOSED	(+)82.13°	IN	KD
3KFA	B	B91	1.22		M.M.	OUT	CLOSED	(+)81.87°	IN	KD
3MS9	A	STI+MS9	1.80		M.M.	OUT	CLOSED (not complete)	(+)83.07°	IN	KD
3MS9	B	STI+MS9	1.80		M.M.	OUT	CLOSED	(+)86.58°	IN	KD
3MSS	A	STI+MS7	1.95		M.M.	OUT	CLOSED (not complete)	(+)83.94°	IN	KD
3MSS	B	STI+MS7	1.95		M.M.	OUT	CLOSED (not complete)	(+)85.34°	IN	KD
3MSS	C	STI+MS7	1.95		M.M.	OUT	CLOSED (not complete)	(+)86.43°	IN	KD
3MSS	D	STI+MS7	1.95		M.M.	OUT	CLOSED (not complete)	(+)89.03°	IN	KD
3OXZ	A	0LI	2.20		M.M.	OUT	CLOSED (not complete)	(+)81.82°	IN	KD
3OY3	A	XY3	1.95	T315I	M.M.	OUT	CLOSED	(+)78.88°	IN	KD
3OY3	B	XY3	1.95	T315I	M.M.	OUT	CLOSED	(+)81.96°	IN	KD
3PYY	A	STI+3YY	1.85		H.S.	OUT	CLOSED	(+)91.69°	IN	KD
3PYY	B	STI+3YY	1.85		H.S.	OUT	N.A.	(+)86.30°	IN	KD
3QRI	A	919	2.10		H.S.	OUT	CLOSED	(+)84.86°	IN	KD
3QRI	B	919	2.10		H.S.	OUT	CLOSED	(+)85.01°	IN	KD
3QRJ	A	919	1.82	T315I	H.S.	OUT	CLOSED	(+)80.26°	IN	KD
3QRJ	B	919	1.82	T315I	H.S.	OUT	CLOSED	(+)86.92°	IN	KD
3QRK	A	9DP	2.30		H.S.	OUT	CLOSED	(+)87.80°	IN	KD
3UE4	A	DB8	2.42		H.S.	OUT	OPEN	(-)83.80°	IN	KD
3UE4	B	DB8	2.42		H.S.	OUT	OPEN	(+)76.68°	IN	KD
4TWP	A	AXI	2.40	T315I	H.S.	IN	OPEN	(+)83.70°	IN	KD
4TWP	B	AXI	2.40	T315I	H.S.	IN	OPEN	(+)78.18°	IN	KD
4WA9	A	AXI	2.20		H.S.	OUT	OPEN	(-)84.31°	N.A.	KD
4WA9	B	AXI	2.20		H.S.	OUT	OPEN	(-)81.76°	IN	KD
4XEY	A	IN1	2.89		H.S.	Intermediate	OPEN	(-)78.66°	OUT	KD
4XEY	B	IN1	2.89		H.S.	Intermediate	OPEN	(-)79.90°	OUT	SH2-KD

Table S1. Properties of available X-ray structures of Abl kinase (ABL1 gene). From left to right: PDB code, chain, ID of ligand co-crystallized, resolution, mutation (if empty is the wild type structure), species (H.M. is Homo Sapiens, while M.M. is Mus Musculus), conformation of DFG motif, conformation of A-loop, dihedral angle value of R381, orientation of the C helix, domains crystallized.

Appendix B

Supporting Material

Automated analysis of binding pocket
dynamics and allosteric networks in
proteins

Supplementary text

System Preparation

All the systems simulated in the present work were prepared by means of AmberTools, version 10 and Gaussian for the ligand partial charge computation.

PNP

The PNP system consists of PNP homotrimeric unit with 9 DADME-ImmH ligand and phosphate ions. The PNP trimer was modeled starting from the PDB code 3K8O using the chains E, Q and Y. The missing first residue (glutamate) in chains E and Q was added manually after superimposing the two chains with chain Y. The tautomerization of H257 was imposed epsilon according to ²¹⁴. DADME-ImmH ligand was retrieved from PDB code 1RSZ and placed around PNP trimer, while phosphate ions were placed in the same positions found in PDB code 1RR6 X-ray structure. Partial charges of DADME-ImmH and phosphate ions were computed using HF/6-31G* theory level, imposing a net charge of +1 for DADME-ImmH. General Amber Force Field (GAFF) was employed to parameterize the ligands and the ions, while partial charges were fitted using the RESP procedure (via Antechamber). O2-P-OH and HO-OH-P angles were modified in order to avoid failures of the SHAKE algorithm.¹⁹⁴ Finally, protein was parameterized with AmberFF99SB-ILDN force field and immersed in TIP3P water box of $\sim 11,000,000 \text{ \AA}^3$, containing nearly 100,000 atoms in total.

Adenosinic Receptor (A2A)

The A2A system was built using the PDB code 3UZC X-ray structure. The missing loop (residues 150-157) was modeled using PDB code 4EIY as template. Alanine was mutated to serine to reproduce the wild type system. The protein was embedded in a POPC membrane bilayer of $75 \times 75 \text{ \AA}^2$ size with a water layer of 30 \AA for each side. The membrane was assembled with CharmmGUI server while the posing and the physical plugging of the protein within the bilayer was done with Membrane Tool present in BiKi LifeSciences.

Abl kinase

For Abl kinase we defined 8 model systems, namely a) SH2-KD_{wt}, b) SH2-KD_{T231R}, c) SH2-KD_{I164E}, d) KD_{out}, e) KD_{in}, f) T315I-KD_{in}, g) Myr/KD_{in} and h) Myr/T315I-KD_{in}. Systems *a*, *b* and *c* were built starting from the PDB code 1OPL chain B X-ray structure which is the KD of Abl in complex with the SH2 domain. For a comprehensive description of the modeling procedures see ref.²⁰⁷ In addition, for the modeling of *b* and *c*, we replaced threonine 231 with arginine and isoleucine 164 with glutamate, respectively. Also the systems *d* and *e*, which are KD alone in DFG-out and DFG-in conformations, respectively, were prepared using PDB code 1OPL chain B and 2F4J structures as reported in ref.²⁰⁷ System *f* was prepared by replacing threonine 315 with isoleucine from system *e*. System *g* was build starting from system *e*, manually adding the myristate molecule at the myristate pocket. To better fit the myristate we employed as reference the PDB code 1OPL chain A. Finally, system *h* was build starting from system *g* and mutating T315 with isoleucine. Point charges of myristate were computed at HF/6-31G* level theory, while the General Amber Force Field (GAFF) was used to parameterize the myristate. All the systems were parameterized with Amber99SB force field.

Molecular dynamics (MD) simulations

Pnp

The system was first equilibrated for 350 ps in NVT ensemble in 5 steps: 150 ps constraining all protein heavy atoms using a harmonic constant $40 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$; 50 ps constraining the protein backbone using a harmonic constant $20 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$; 50 ps where the protein backbone is constrained at $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and 50 ps with a constraint of $1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for the protein backbone. Then the system underwent to other 350 ps MD simulations in NPT ensemble at 1 bar. For the production phase we run about 1000 ns of simulations in NVT ensemble using ACEMD engine. The equation of motion was integrated every 2 fs. Bond involving hydrogen atoms were restrained to their equilibrium length with the SHAKE algorithm. A short-range nonbonded cutoff of 9 \AA was applied, whereas long-range electrostatics was treated with the Particle Mesh Ewald method. Temperature was maintained at 300 K using Langevin thermostat and damping of 0.1 ps.

Adenosinic Receptor (A2A)

The system was equilibrated for 300 ps in NVT ensemble employing 3 steps of 100 ps each at 100, 200 and 300 K. In the first two steps a harmonic constraint of 1000 kJ/mol/Å² was applied at the protein backbone. Then, 1 ns at NPT ensemble was run to reach pressure equilibrium. The production was run for 100 ns in NPT ensemble. A short-range nonbonded cutoff of 11 Å was applied, whereas long-range electrostatics was treated with the PME method. Bussi-Parrinello thermostat was employed and to constraint all bonds LINCS algorithm was employed. Pressure was kept anisotropic (X,Y plane). MD simulations were performed with Gromacs 4.6.1 engine.

Abl kinase

First the systems underwent to a heating phase, in which the systems reached 300 K in 100 ps in NVT ensemble, keeping the C α atoms fixed in their original positions, using an harmonic force constant of 1000 kJ/mol/Å². Then the pressure was equilibrated running 5 ns of NPT simulation reaching 1 bar and maintaining C α restrained. Production phase was performed in NPT ensemble. Bond involving hydrogen atoms were restrained with the LINCS algorithm. A short-range nonbonded cutoff of 9 Å was applied, whereas long-range electrostatics was treated with the PME method. The equations of motion was integrated every 2 fs. The Parrinello-Rahaman barostat and the velocity-rescaling thermostat were employed with a relaxation time τ of 2 ps and 0.1 ps, respectively. Coordinates of the systems were collected every 50 ps for each run. All the simulations were performed with Gromacs 4.6 engine. Overall, we collected \sim 1.8 μ s for SH2-KD_{wt}, \sim 1.5 μ s for SH2-KD_{T231R}, \sim 1.3 μ s for SH2-KD_{I164E}, \sim 1.3 μ s for KD_{in}, \sim 1.3 μ s for KD_{out}, \sim 1.0 μ s for T315I-KD_{in}, \sim 0.8 μ s for Myr/T315I-KD_{in} and \sim 1.0 μ s for Myr/KD_{in}.

Supplementary figures

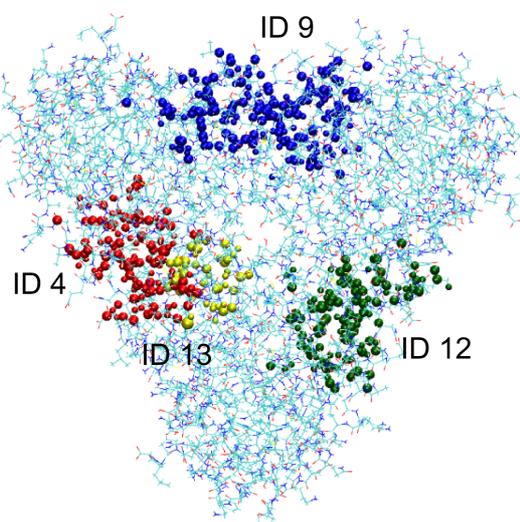


Figure S1. Representation of atoms that underlie the three orthosteric pockets in PNP. The first and the second orthosteric pockets are represented by pID 9 (blue spheres) and pID 12 (green spheres). The union of pID4 and pID 13 (red and yellow, respectively) represent the third orthosteric pocket.

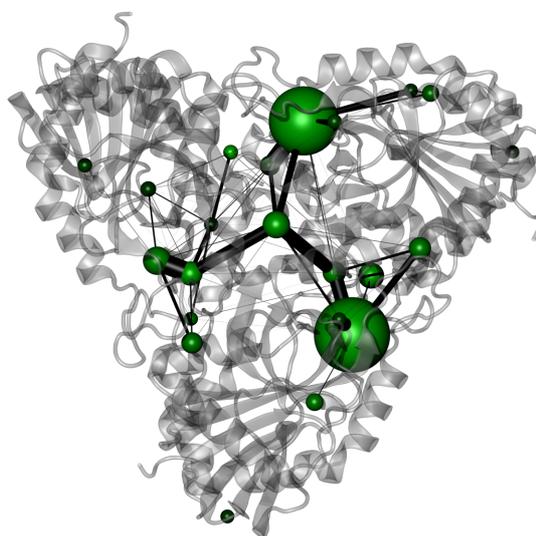


Figure S2. Pocket network on PNP system. Each pocket (i.e. node of the network) is represented as a green sphere, while the black line (i.e. edge of the network) represent the communication between pockets. Only pockets with a persistency over the simulation time higher that 30% are represented.

Supplementary tables

Table S1. Time persistency of residues that define the ATP pocket in KD_{out} and KD_{in} .

Residue	KD_{out}			KD_{in}	
	pID 5	pID 28	pID 3	pID 5	pID 3
L267	83,4	0	0	90,5	0
G268	0	0	0	61	0
G269	0	0	0	65,4	0
G270	0	0	0	60,9	0
Q271	0	0	0	72,8	0
Y272	49,4	0	0	54,1	0
G273	25,8	0	0	0	0
E274	21,8	0	0	0	0
V275	90,7	0	0	97,2	0
A288	90,5	0	0	97,8	0
V289	20,3	0	0	49,7	0
K290	87,9	33,8	0	97,8	0
T291	0	27	0	0	0
L292	0	29,2	0	0	0
K293	0	31,6	0	0	0
E301	0	20,4	0	0	43,4
F302	25,4	34,1	0	0	21,1
K304	0	0	0	0	34,3
E305	35,3	31,5	33,4	83	56,6
A306	26,4	0	0	0	0
V307	0	0	32,6	0	50,8
M308	42,9	0	22,6	81,4	0
V317	85,9	0	0	91,6	0
L319	26	0	0	0	0
I332	50,8	20,7	0	74,4	0
T334	89,3	0	0	97,8	0
E335	74,8	0	0	69,2	0
F336	85	0	0	77	0
N341	49,1	0	0	80,2	0
P378	0	0	21,6	0	21,3
R381	0	0	0	0	28,6
D382	0	0	0	56,2	0
R386	49,2	0	0	82	0
N387	37,4	0	0	83,8	0
C388	0	0	0	22,4	0
L389	89,3	0	0	93,3	0
A399	88,7	0	0	97,8	0
D400	88,1	0	0	97,9	0

F401	90,5	0	0	64,3	48,7
G402	0	0	0	32	58,8
L403	29,7	0	26	33,7	52,1
S404	0	0	29,3	0	55,6
R405	0	0	24,5	0	61,4

Bibliography

1. Ribeiro, A. A. S. T. & Ortiz, V. A Chemical Perspective on Allostery. *Chem. Rev.* **116**, 6488–6502 (2016).
2. Christopoulos, A. Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nat. Rev. Drug Discov.* **1**, 198–210 (2002).
3. Nussinov, R. & Tsai, C. J. Unraveling structural mechanisms of allosteric drug action. *Trends in Pharmacological Sciences* **35**, 256–264 (2014).
4. Wagner, J. R. *et al.* Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem. Rev.* **116**, 6370–6390 (2016).
5. van Westen, G. J. P., Gaulton, A., Overington, J. P., Augustin, T. & Zeileis, A. Chemical, Target, and Bioactive Properties of Allosteric Modulation. *PLoS Comput. Biol.* **10**, e1003559 (2014).
6. Henrich, S. *et al.* Computational approaches to identifying and characterizing protein binding sites for ligand design. *J. Mol. Recognit.* **23**, 209–219 (2009).
7. Tsai, C.-J. & Nussinov, R. A Unified View of ‘How Allostery Works’. *PLoS Comput. Biol.* **10**, e1003394 (2014).
8. Salesse, S. & Verfaillie, C. M. BCR/ABL: from molecular mechanisms of leukemia induction to treatment of chronic myelogenous leukemia. *Oncogene* **21**, 8547–8559 (2002).
9. Changeux, J.-P. Allosteric Mechanisms of Signal Transduction. *Science (80-.)*. **308**, 1424–1428 (2005).
10. Philips, S. J. *et al.* Allosteric transcriptional regulation via changes in the overall topology of the core promoter. *Science (80-.)*. **349**, 877–881 (2015).
11. Bray, D. & Duke, T. Conformational Spread: The Propagation of Allosteric States in Large Multiprotein Complexes. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 53–73 (2004).
12. Link, H., Christodoulou, D. & Sauer, U. Advancing metabolic models with kinetic information. *Curr. Opin. Biotechnol.* **29**, 8–14 (2014).
13. Nussinov, R. & Tsai, C.-J. Allostery in Disease and in Drug Discovery. *Cell* **153**, 293–305 (2013).
14. Schueler-Furman, O. & Wodak, S. J. Computational approaches to investigating allostery. *Curr. Opin. Struct. Biol.* **41**, 159–171 (2016).
15. Nussinov, R. Introduction to Protein Ensembles and Allostery. *Chem. Rev.* **116**, 6263–6266 (2016).
16. Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).
17. Koshland, D. E., Némethy, G. & Filmer, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **5**, 365–385 (1966).
18. Tsai, C.-J. *et al.* Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Mol. Biosyst.* **5**, 207 (2009).
19. Cuendet, M. A., Weinstein, H. & LeVine, M. V. The Allostery Landscape: Quantifying Thermodynamic Couplings in Biomolecular Systems. *J. Chem. Theory Comput.* **12**, 5758–5767 (2016).
20. Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. *Nature* **508**, 331–339 (2014).
21. Lu, S., Li, S. & Zhang, J. Harnessing Allostery: A Novel Approach to Drug Discovery. *Med. Res. Rev.* **34**, 1242–1285 (2014).
22. Kornev, A. P. & Taylor, S. S. Dynamics-Driven Allostery in Protein Kinases. *Trends Biochem. Sci.* **40**, 628–647 (2015).
23. Lockless, S. W. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science (80-.)*. **286**, 295–299 (2009).
24. Süel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* **10**, 59–69 (2003).
25. Perutz, M. F., Muirhead, H., Cox, J. M. & Goaman, L. C. Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model. *Nature* **219**, 131–139 (1968).

26. Ye, L., Van Eps, N., Zimmer, M., Ernst, O. P. & Scott Prosser, R. Activation of the A2A adenosine G-protein-coupled receptor by conformational selection. *Nature* **533**, 265–268 (2016).
27. Ramanoudjame, G., Du, M., Mankiewicz, K. A. & Jayaraman, V. Allosteric mechanism in AMPA receptors: a FRET-based investigation of conformational changes. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10473–8 (2006).
28. Chandramohan, A. *et al.* Predicting Allosteric Effects from Orthosteric Binding in Hsp90-Ligand Interactions: Implications for Fragment-Based Drug Design. *PLOS Comput. Biol.* **12**, e1004840 (2016).
29. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat Struct Mol Biol* **9**, 646–652 (2002).
30. de Juan, D., Pazos, F. & Valencia, A. Emerging methods in protein co-evolution. *Nat. Rev. Genet.* **14**, 249–261 (2013).
31. Avila-Herrera, A. & Pollard, K. S. Coevolutionary analyses require phylogenetically deep alignments and better null models to accurately detect inter-protein contacts within and between species. *BMC Bioinformatics* **16**, 268 (2015).
32. Ferreira, D. U., Komives, E. A. & Wolynes, P. G. Frustration in biomolecules. *Q. Rev. Biophys.* **47**, 285–363 (2014).
33. Ferreira, D. U., Hegler, J. A., Komives, E. A. & Wolynes, P. G. On the role of frustration in the energy landscapes of allosteric proteins. *Proc. Natl. Acad. Sci.* **108**, 3499–3503 (2011).
34. Weinkam, P., Pons, J. & Sali, A. Structure-based model of allostery predicts coupling between distant sites. *Proc. Natl. Acad. Sci.* **109**, 4875–4880 (2012).
35. Skjaerven, L., Hollup, S. M. & Reuter, N. Normal mode analysis for proteins. *J. Mol. Struct. THEOCHEM* **898**, 42–48 (2009).
36. Gaillard, T. *et al.* Comparative Normal Mode Analysis of LFA-1 Integrin I-domains. *J. Mol. Biol.* **374**, 231–249 (2007).
37. Greener, J. G. & Sternberg, M. J. AlloPred: prediction of allosteric pockets on proteins using normal mode perturbation analysis. *BMC Bioinformatics* **16**, 335 (2015).
38. Ming, D. & Wall, M. E. Allostery in a Coarse-Grained Model of Protein Dynamics. *Phys. Rev. Lett.* **95**, 198103 (2005).
39. Bahar, I., Chennubhotla, C. & Tobi, D. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr. Opin. Struct. Biol.* **17**, 633–640 (2007).
40. Tama, F., Valle, M., Frank, J. & Brooks, C. L. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9319–23 (2003).
41. Tama, F. & Brooks, C. L. The Mechanism and Pathway of pH Induced Swelling in Cowpea Chlorotic Mottle Virus. *J. Mol. Biol.* **318**, 733–747 (2002).
42. Rader, A. J., Vlad, D. H. & Bahar, I. Maturation Dynamics of Bacteriophage HK97 Capsid. *Structure* **13**, 413–421 (2005).
43. McClendon, C. L., Friedland, G., Mobley, D. L., Amirkhani, H. & Jacobson, M. P. Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles. *J. Chem. Theory Comput.* **5**, 2486–2502 (2009).
44. Bradley, M. J., Chivers, P. T. & Baker, N. A. Molecular Dynamics Simulation of the Escherichia coli NikR Protein: Equilibrium Conformational Fluctuations Reveal Interdomain Allosteric Communication Pathways. *J. Mol. Biol.* **378**, 1155–1173 (2008).
45. Rehn, A. *et al.* Allosteric Regulation Points Control the Conformational Dynamics of the Molecular Chaperone Hsp90. *J. Mol. Biol.* **428**, 4559–4571 (2016).
46. Rivalta, I. *et al.* Allosteric pathways in imidazole glycerol phosphate synthase. *Proc. Natl. Acad. Sci.* **109**, E1428–E1436 (2012).
47. Pérot, S., Sperandio, O., Miteva, M. A., Camproux, A.-C. & Villoutreix, B. O. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov. Today* **15**, 656–667 (2010).
48. Krone, M. *et al.* Visual Analysis of Biomolecular Cavities: State of the Art. *Comput. Graph. Forum* **35**, 527–551 (2016).
49. Schmidtke, P., Bidon-chanal, A., Luque, F. J. & Barril, X. MDpocket: Open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **27**, 3276–3285 (2011).
50. Manning, G. The Protein Kinase Complement of the Human Genome. *Science (80-.)*. **298**, 1912–1934 (2002).
51. Cohen, P. The origins of protein phosphorylation. *Nat. Cell Biol.* **4**, E127–E130 (2002).

52. Johnson, L. N. & Lewis, R. J. Structural basis for control by phosphorylation. *Chem. Rev.* **101**, 2209–2242 (2001).
53. Kuriyan, J. & Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **450**, 983–990 (2007).
54. Wu, P., Nielsen, T. E. & Clausen, M. H. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol. Sci.* **36**, 422–439 (2015).
55. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275–282 (2002).
56. Johnson, L. N., Noble, M. E. M. & Owen, D. J. Active and inactive protein kinases: Structural basis for regulation. *Cell* **85**, 149–158 (1996).
57. Kornev, A. P., Haste, N. M., Taylor, S. S. & Eyck, L. F. Ten. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 17783–17788 (2006).
58. Kornev, A. P., Taylor, S. S. & Ten Eyck, L. F. A helix scaffold for the assembly of active protein kinases. *Proc. Natl. Acad. Sci.* **105**, 14377–14382 (2008).
59. Taylor, S. S., Ilouz, R., Zhang, P. & Kornev, A. P. Assembly of allosteric macromolecular switches: lessons from PKA. *Nat. Rev. Mol. Cell Biol.* **13**, 646–658 (2012).
60. Hu, S.-H. *et al.* Insights into autoregulation from the crystal structure of twitchin kinase. *Nature* **369**, 581–584 (1994).
61. Wilmanns, M. *et al.* Structural basis for activation of the titin kinase domain during myofibrillogenesis. *Nature* **395**, 863–869 (1998).
62. Jura, N. *et al.* Catalytic Control in the EGF Receptor and Its Connection to General Kinase Regulatory Mechanisms. *Mol. Cell* **42**, 9–22 (2011).
63. Khatri, A., Wang, J. & Pendergast, A. M. Multifunctional Abl kinases in health and disease. *J. Cell Sci.* **129**, 9–16 (2016).
64. Wong, S. & Witte, O. N. The BCR-ABL Story: Bench to Bedside and Back. *Annu. Rev. Immunol.* **22**, 247–306 (2004).
65. Eide, C. A. & O'Hare, T. Chronic Myeloid Leukemia: Advances in Understanding Disease Biology and Mechanisms of Resistance to Tyrosine Kinase Inhibitors. *Curr. Hematol. Malig. Rep.* **10**, 158–166 (2015).
66. Hantschel, O., Grebien, F. & Superti-Furga, G. The Growing Arsenal of ATP-Competitive and Allosteric Inhibitors of BCR-ABL. *Cancer Res.* **72**, 4890–4895 (2012).
67. Hantschel, O. & Superti-Furga, G. Regulation of the c-Abl and Bcr-Abl tyrosine kinases. *Nat. Rev. Mol. Cell Biol.* **5**, 33–44 (2004).
68. Panjarian, S., Iacob, R. E., Chen, S., Engen, J. R. & Smithgall, T. E. Structure and dynamic regulation of abl kinases. *J. Biol. Chem.* **288**, 5443–5450 (2013).
69. Nagar, B. *et al.* Organization of the SH3-SH2 Unit in Active and Inactive Forms of the c-Abl Tyrosine Kinase. *Mol. Cell* **21**, 787–798 (2006).
70. Lorenz, S., Deng, P., Hantschel, O., Superti-Furga, G. & Kuriyan, J. Crystal structure of an SH2-kinase construct of c-Abl and effect of the SH2 domain on kinase activity. *Biochem. J.* **468**, 283–291 (2015).
71. Pluk, H., Dorey, K. & Superti-Furga, G. Autoinhibition of c-Abl. *Cell* **108**, 247–259 (2002).
72. Nagar, B. *et al.* Structural Basis for the Autoinhibition of c-Abl Tyrosine Kinase. *Cell* **112**, 859–871 (2003).
73. Pisabarro, M. T., Serrano, L. & Wilmanns, M. Crystal structure of the abl-SH3 domain complexed with a designed high-affinity peptide ligand: implications for SH3-ligand interactions. *J. Mol. Biol.* **281**, 513–521 (1998).
74. Franz, W. M., Berger, P. & Wang, J. Y. Deletion of an N-terminal regulatory domain of the c-abl tyrosine kinase activates its oncogenic potential. *EMBO J* **8**, 137–147 (1989).
75. Wojcik, J. *et al.* Allosteric Inhibition of Bcr-Abl Kinase by High Affinity Monobody Inhibitors Directed to the Src Homology 2 (SH2)-Kinase Interface. *J. Biol. Chem.* **291**, 8836–8847 (2016).
76. Grebien, F. *et al.* Targeting the SH2-Kinase Interface in Bcr-Abl Inhibits Leukemogenesis. *Cell* **147**, 306–319 (2011).
77. Zuccotto, F., Ardini, E., Casale, E. & Angiolini, M. Through the 'Gatekeeper Door': Exploiting the Active Kinase Conformation. *J. Med. Chem.* **53**, 2681–2694 (2010).
78. Reddy, E. P. & Aggarwal, A. K. The Ins and Outs of Bcr-Abl Inhibition. *Genes Cancer* **3**, 447–454 (2012).
79. Agarwal, S. & Mehrotra, R. An overview of Molecular Docking. *JSM Chem* **4**, 1024 (2016).
80. Meng, X.-Y., Zhang, H.-X., Mezei, M. & Cui, M. Molecular docking: a powerful approach

- for structure-based drug discovery. *Curr. Comput. Aided. Drug Des.* **7**, 146–157 (2011).
81. Taylor, R. D., Jewsbury, P. J. & Essex, J. W. A review of protein-small molecule docking methods. *J. Comput. Aided. Mol. Des.* **16**, 151–166 (2002).
 82. Halperin, I., Ma, B., Wolfson, H. & Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins Struct. Funct. Genet.* **47**, 409–443 (2002).
 83. Neves, M. A. C., Totrov, M. & Abagyan, R. Docking and scoring with ICM: The benchmarking results and strategies for improvement. *J. Comput. Aided. Mol. Des.* **26**, 675–686 (2012).
 84. Friesner, R. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **47**, 1739–1749 (2004).
 85. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748 (1997).
 86. Kramer, B., Rarey, M. & Lengauer, T. Evaluation of the FlexX incremental construction algorithm for protein-ligand docking. *Proteins Struct. Funct. Genet.* **37**, 228–241 (1999).
 87. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161**, 269–288 (1982).
 88. Kollman, P. Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **93**, 2395–2417 (1993).
 89. Böhm, H. J. Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput. Aided. Mol. Des.* **12**, 309–323 (1998).
 90. Muegge, I. & Martin, Y. C. A General and Fast Scoring Function for Protein-Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **42**, 791–804 (1999).
 91. Sun, H. *et al.* Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Phys. Chem. Chem. Phys.* **16**, 22035–22045 (2014).
 92. Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **3**, 935–949 (2004).
 93. Norberg, J. & Nilsson, L. Molecular dynamics applied to nucleic acids. *Acc. Chem. Res.* **35**, 465–472 (2002).
 94. Jones, J. E. On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **106**, 463–477 (1924).
 95. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
 96. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
 97. Schmid, N., Christ, C. D., Christen, M., Eichenberger, A. P. & van Gunsteren, W. F. Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation. *Comput. Phys. Commun.* **183**, 890–903 (2012).
 98. Durrant, J. D. & McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **9**, 71 (2011).
 99. Sborgi, L. *et al.* Interaction Networks in Protein Folding via Atomic-Resolution Experiments and Long-Time-Scale Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **137**, 6506–6516 (2015).
 100. Shan, Y. *et al.* How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **133**, 9181–9183 (2011).
 101. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci.* **99**, 12562–12566 (2002).
 102. Van Speybroeck, V. *et al.* First principle chemical kinetics in zeolites: the methanol-to-olefin process as a case study. *Chem. Soc. Rev.* **43**, 7326–7357 (2014).
 103. Branduardi, D., Gervasio, F. L., Cavalli, A., Recanatini, M. & Parrinello, M. The role of the peripheral anionic site and cation- π interactions in the ligand penetration of the human AChE gorge. *J. Am. Chem. Soc.* **127**, 9147–9155 (2005).
 104. Bussi, G., Laio, A. & Parrinello, M. Equilibrium free energies from nonequilibrium metadynamics. *Phys. Rev. Lett.* **96**, 90601 (2006).
 105. Laio, A., Rodriguez-Fortea, A., Gervasio, F. L., Ceccarelli, M. & Parrinello, M. Assessing the accuracy of metadynamics. *J. Phys. Chem. B* **109**, 6714–6721 (2005).
 106. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: A smoothly

- converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 20603 (2008).
107. Rahman, J. A. & Tully, J. C. Puddle-skimming: An efficient sampling of multidimensional configuration space. *J. Chem. Phys.* **116**, 8750–8760 (2002).
 108. Hamelberg, D., Mongan, J. & McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **120**, 11919–11929 (2004).
 109. De Oliveira, C. A. F., Hamelberg, D. & McCammon, J. A. On the application of accelerated molecular dynamics to liquid water simulations. *J. Phys. Chem. B* **110**, 22695–22701 (2006).
 110. Miao, Y., Feher, V. A. & McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **11**, 3584–3595 (2015).
 111. Van Etten, R. A. Cycling, stressed-out and nervous: cellular functions of c-Abl. *Trends Cell Biol* **9**, 179–186 (1999).
 112. Sirvent, A., Benistant, C. & Roche, S. Cytoplasmic signalling by the c-Abl tyrosine kinase in normal and cancer cells. *Biol. Cell* **100**, 617–631 (2008).
 113. Deininger, M., Buchdunger, E. & Druker, B. J. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood* **105**, 2640–2653 (2005).
 114. Zhang, J. *et al.* Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. *Nature* **463**, 501–506 (2010).
 115. Lamontanara, A. J., Gencer, E. B., Kuzyk, O. & Hantschel, O. Mechanisms of resistance to BCR-ABL and other kinase inhibitors. *Biochim. Biophys. Acta - Proteins Proteomics* **1834**, 1449–1459 (2013).
 116. Cavalli, A., De Vivo, M. & Recanatini, M. Density functional study of the enzymatic reaction catalyzed by a cyclin-dependent kinase. *Chem. Commun. (Camb)*. 1308–9 (2003).
 117. De Vivo, M., Cavalli, A., Carloni, P. & Recanatini, M. Computational study of the phosphoryl transfer catalyzed by a cyclin-dependent kinase. *Chemistry (Easton)*. **13**, 8437–8444 (2007).
 118. Foda, Z. H., Shan, Y., Kim, E. T., Shaw, D. E. & Seeliger, M. A. A dynamically coupled allosteric network underlies binding cooperativity in Src kinase. *Nat. Commun.* **6**, 5939 (2015).
 119. Skora, L., Mestan, J., Fabbro, D., Jahnke, W. & Grzesiek, S. NMR reveals the allosteric opening and closing of Abelson tyrosine kinase by ATP-site and myristoyl pocket inhibitors. *Proc. Natl. Acad. Sci.* **110**, E4437–E4445 (2013).
 120. De Vivo, M., Cavalli, A., Bottegoni, G., Carloni, P. & Recanatini, M. Role of phosphorylated Thr160 for the activation of the CDK2/Cyclin A complex. *Proteins* **62**, 89–98 (2006).
 121. Meng, Y. & Roux, B. Locking the Active Conformation of c-Src Kinase through the Phosphorylation of the Activation Loop. *J. Mol. Biol.* **426**, 423–435 (2014).
 122. Tanis, K. Q., Veach, D., Duewel, H. S., Bornmann, W. G. & Koleske, A. J. Two Distinct Phosphorylation Pathways Have Additive Effects on Abl Family Kinase Activation. *Mol. Cell. Biol.* **23**, 3884–3896 (2003).
 123. Brasher, B. B. & Van Etten, R. A. c-Abl Has High Intrinsic Tyrosine Kinase Activity That Is Stimulated by Mutation of the Src Homology 3 Domain and by Autophosphorylation at Two Distinct Regulatory Tyrosines. *J. Biol. Chem.* **275**, 35631–35637 (2000).
 124. Dorey, K. *et al.* Phosphorylation and structure-based functional studies reveal a positive and a negative role for the activation loop of the c-Abl tyrosine kinase. *Oncogene* **20**, 8075–8084 (2001).
 125. Shan, Y. *et al.* A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc. Natl. Acad. Sci.* **106**, 139–144 (2009).
 126. Oruganty, K., Talathi, N. S., Wood, Z. A. & Kannan, N. Identification of a hidden strain switch provides clues to an ancient structural mechanism in protein kinases. *Proc. Natl. Acad. Sci.* **110**, 924–929 (2013).
 127. Filippakopoulos, P. *et al.* Structural Coupling of SH2-Kinase Domains Links Fes and Abl Substrate Recognition and Kinase Activation. *Cell* **134**, 793–803 (2008).
 128. Lamontanara, A. J., Georgeon, S., Tria, G., Svergun, D. I. & Hantschel, O. The SH2 domain of Abl kinases regulates kinase autophosphorylation by controlling activation loop accessibility. *Nat. Commun.* **5**, 5470 (2014).
 129. Dölker, N. *et al.* The SH2 domain regulates c-Abl kinase activation by a cyclin-like mechanism and remodulation of the hinge motion. *PLoS Comput. Biol.* **10**, e1003863 (2014).
 130. Tse, A. & Verkhivker, G. M. Molecular Dynamics Simulations and Structural Network Analysis of c-Abl and c-Src Kinase Core Proteins: Capturing Allosteric Mechanisms and Communication Pathways from Residue Centrality. *J. Chem. Inf. Model.* **55**, 1645–1662

- (2015).
131. De Vivo, M. *et al.* Cyclin-dependent kinases: bridging their structure and function through computations. *Future Med. Chem.* **3**, 1551–1559 (2011).
 132. De Vivo, M., Masetti, M., Bottegoni, G. & Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **59**, 4035–4061 (2016).
 133. Nolen, B., Taylor, S. & Ghosh, G. Regulation of Protein Kinases: Controlling Activity through Activation Segment Conformation. *Mol. Cell* **15**, 661–675 (2004).
 134. Cox, S., Radzio-Andzelm, E. & Taylor, S. S. Domain movements in protein kinases. *Curr. Opin. Struct. Biol.* **4**, 893–901 (1994).
 135. Dixit, A. & Verkhivker, G. M. Computational Modeling of Allosteric Communication Reveals Organizing Principles of Mutation-Induced Signaling in ABL and EGFR Kinases. *PLoS Comput. Biol.* **7**, e1002179 (2011).
 136. Lou, H. & Cukier, R. I. Molecular dynamics of apo-adenylate kinase: A principal component analysis. *J. Phys. Chem. B* **110**, 12796–12808 (2006).
 137. Cheng, Y., Zhang, Y. & McCammon, J. A. How does activation loop phosphorylation modulate catalytic activity in the cAMP-dependent protein kinase: a theoretical study. *Protein Sci.* **15**, 672–683 (2006).
 138. Lu, B., Wong, C. F. & McCammon, J. A. Release of ADP from the catalytic subunit of protein kinase A: A molecular dynamics simulation study. *Protein Sci.* **14**, 159–168 (2005).
 139. Zhang, L. *et al.* Functional Role of Histidine in the Conserved His-x-Asp Motif in the Catalytic Core of Protein Kinases. *Sci. Rep.* **5**, 10115 (2015).
 140. Nagar, B. *et al.* Crystal Structures of the Kinase Domain of c-Abl in Complex with the Small Molecule Inhibitors PD173955 and Imatinib (STI-571). *Cancer Res.* **62**, 4236–4243 (2002).
 141. Cowan-Jacob, S. W. *et al.* Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63**, 80–93 (2007).
 142. Jahnke, W. *et al.* Binding or Bending: Distinction of Allosteric Abl Kinase Agonists from Antagonists by an NMR-Based Conformational Assay. *J. Am. Chem. Soc.* **132**, 7043–7048 (2010).
 143. Yang, J. *et al.* Discovery and Characterization of a Cell-Permeable, Small-Molecule c-Abl Kinase Activator that Binds to the Myristoyl Binding Site. *Chem. Biol.* **18**, 177–186 (2011).
 144. Schindler, T. *et al.* Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science (80-.)*. **289**, 1938–1942 (2000).
 145. Vajpai, N. *et al.* Solution Conformations and Dynamics of ABL Kinase-Inhibitor Complexes Determined by NMR Substantiate the Different Binding Modes of Imatinib/Nilotinib and Dasatinib. *J. Biol. Chem.* **283**, 18292–18302 (2008).
 146. Ensing, B., De Vivo, M., Liu, Z., Moore, P. & Klein, M. L. Metadynamics as a Tool for Exploring Free Energy Landscapes of Chemical Reactions. *Acc. Chem. Res.* **39**, 73–81 (2006).
 147. Strong, T. C., Kaur, G. & Thomas, J. H. Mutations in the catalytic loop HRD motif alter the activity and function of Drosophila Src64. *PLoS One* **6**, e28100 (2011).
 148. Kannan, N. & Neuwald, A. F. Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component? *J. Mol. Biol.* **351**, 956–972 (2005).
 149. Walter, A. O., Seghezzi, W., Korver, W., Sheung, J. & Lees, E. The mitotic serine/threonine kinase Aurora2/AIK is regulated by phosphorylation and degradation. *Oncogene* **19**, (2000).
 150. Littlepage, L. E. *et al.* Identification of phosphorylated residues that affect the activity of the mitotic kinase Aurora-A. *Proc. Natl. Acad. Sci.* **99**, 15440–15445 (2002).
 151. Roskoski, R. Src protein–tyrosine kinase structure and regulation. *Biochem. Biophys. Res. Commun.* **324**, 1155–1164 (2004).
 152. Porter, M., Schindler, T., Kuriyan, J. & Miller, W. T. Reciprocal Regulation of Hck Activity by Phosphorylation of Tyr527 and Tyr416: effect of introducing a high affinity intramolecular SH2 ligand. *J. Biol. Chem.* **275**, 2721–2726 (2000).
 153. Bayliss, R., Sardon, T., Vernos, I. & Conti, E. Structural Basis of Aurora-A Activation by TPX2 at the Mitotic Spindle. *Mol. Cell* **12**, 851–862 (2003).
 154. Clark, M. A. *et al.* Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat Chem Biol* **5**, 647–654 (2009).
 155. Azam, M., Seeliger, M. A., Gray, N. S., Kuriyan, J. & Daley, G. Q. Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nat Struct Mol Biol* **15**, 1109–1118 (2008).
 156. Sabat, M. *et al.* The development of 2-benzimidazole substituted pyrimidine based inhibitors of lymphocyte specific kinase (Lck). *Bioorg. Med. Chem. Lett.* **16**, 5973–5977 (2006).

157. Fancelli, D. *et al.* 1,4,5,6-Tetrahydropyrrolo[3,4-c]pyrazoles: Identification of a Potent Aurora Kinase Inhibitor with a Favorable Antitumor Kinase Inhibition Profile. *J. Med. Chem.* **49**, 7247–7251 (2006).
158. Xu, W., Harrison, S. C. & Eck, M. J. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* **385**, 595–602 (1997).
159. Sicheri, F., Moarefi, I. & Kuriyan, J. Crystal structure of the Src family tyrosine kinase Hck. *Nature* **385**, 602–609 (1997).
160. De Vivo, M. Bridging quantum mechanics and structure-based drug design. *Front. Biosci. (Landmark Ed.)* **16**, 1619–33 (2011).
161. Fallacara, A. L., Tintori, C., Radi, M., Schenone, S. & Botta, M. Insight into the Allosteric Inhibition of Abl Kinase. *J. Chem. Inf. Model.* **54**, 1325–1338 (2014).
162. Tintori, C. *et al.* Studies on the ATP binding site of Fyn kinase for the identification of new inhibitors and their evaluation as potential agents against tauopathies and tumors. *J. Med. Chem.* **58**, 4590–4609 (2015).
163. Young, M. A. *et al.* Structure of the Kinase Domain of an Imatinib-Resistant Abl Mutant in Complex with the Aurora Kinase Inhibitor VX-680. *Cancer Res.* **66**, 1007–1014 (2006).
164. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
165. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinforma.* **65**, 712–725 (2006).
166. Amadei, A., Linssen, A. B. M., de Groot, B. L., van Aalten, D. M. F. & Berendsen, H. J. C. An Efficient Method for Sampling the Essential Subspace of Proteins. *J. Biomol. Struct. Dyn.* **13**, 615–625 (1996).
167. Oruganty, K. & Kannan, N. Design principles underpinning the regulatory diversity of protein kinases. *Philos. Trans. R. Soc. London B Biol. Sci.* **367**, 2529–2539 (2012).
168. Raiteri, P., Laio, A., Gervasio, F. L., Micheletti, C. & Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem. B* **110**, 3533–3539 (2006).
169. Crespo, Y., Marinelli, F., Pietrucci, F. & Laio, A. Metadynamics convergence law in a multidimensional system. *Phys. Rev. E* **81**, 55701 (2010).
170. Baftizadeh, F., Cossio, P., Pietrucci, F. & Laio, A. Protein Folding and Ligand-Enzyme Binding from Bias-Exchange Metadynamics Simulations. *Curr. Phys. Chem.* **2**, 79–91 (2012).
171. McGovern, M. & de Pablo, J. A boundary correction algorithm for metadynamics in multiple dimensions. *J. Chem. Phys.* **139**, 84102 (2013).
172. Dama, J. F., Hocky, G. M., Sun, R. & Voth, G. A. Exploring Valleys without Climbing Every Peak: More Efficient and Forgiving Metabasin Metadynamics via Robust On-the-Fly Bias Domain Restriction. *J. Chem. Theory Comput.* **11**, 5638–5650 (2015).
173. Yaffe, E., Fishelovitch, D., Wolfson, H. J., Halperin, D. & Nussinov, R. MolAxis: Efficient and accurate identification of channels in macromolecules. *Proteins Struct. Funct. Bioinforma.* **73**, 72–86 (2008).
174. Petřek, M., Košinová, P., Koča, J. & Otyepka, M. MOLE: A Voronoi Diagram-Based Explorer of Molecular Channels, Pores, and Tunnels. *Structure* **15**, 1357–1363 (2007).
175. Levitt, D. G. & Banaszak, L. J. POCKET: A computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *J. Mol. Graph.* **10**, 229–234 (1992).
176. An, J. Pocketome via Comprehensive Identification and Classification of Ligand Binding Envelopes. *Mol. Cell. Proteomics* **4**, 752–761 (2005).
177. Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. P. HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360 (1996).
178. Laskowski, R. A. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graph.* **13**, 323–330 (1995).
179. Decherchi, S. & Rocchia, W. A general and robust ray-casting-based algorithm for triangulating surfaces at the nanoscale. *PLoS One* **8**, e59744 (2013).
180. Pérot, S., Sperandio, O., Miteva, M. a., Camproux, A.-C. & Villoutreix, B. O. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov. Today* **15**, 656–667 (2010).
181. Craig, I. R., Pflieger, C., Gohlke, H., Essex, J. W. & Spiegel, K. Pocket-space maps to identify novel binding-site conformations in proteins. *J. Chem. Inf. Model.* **51**, 2666–2679 (2011).
182. Laurent, B. *et al.* Epock: rapid analysis of protein pocket dynamics. *Bioinformatics* **31**, 1478–

- 1480 (2015).
183. Paramo, T., East, A., Garzón, D., Ulmschneider, M. B. & Bond, P. J. Efficient Characterization of Protein Cavities within Molecular Simulation Trajectories: trj_cavity. *J. Chem. Theory Comput.* **10**, 2151–2164 (2014).
 184. Kokh, D. B. *et al.* TRAPP: A Tool for Analysis of Transient Binding Pockets in Proteins. *J. Chem. Inf. Model.* **53**, 1235–1252 (2013).
 185. Ashford, P. *et al.* Visualisation of variable binding pockets on protein surfaces by probabilistic analysis of related structure sets. *BMC Bioinformatics* **13**, 39 (2012).
 186. Eyrisch, S. & Helms, V. Transient pockets on protein surfaces involved in protein-protein interaction. *J. Med. Chem.* **50**, 3457–3464 (2007).
 187. Panjkovich, A. & Daura, X. Exploiting protein flexibility to predict the location of allosteric sites. *BMC Bioinformatics* **13**, 273 (2012).
 188. Demerdash, O. N. A., Daily, M. D. & Mitchell, J. C. Structure-Based Predictive Models for Allosteric Hot Spots. *PLoS Comput. Biol.* **5**, e1000531 (2009).
 189. Ealick, S. E. *et al.* Three-dimensional structure of human erythrocytic purine nucleoside phosphorylase at 3.2 Å resolution. *J. Biol. Chem.* **265**, 1812–1820 (1990).
 190. Williams, S. R., Gekeler, V., McIvor, R. S. & Martin, D. W. A human purine nucleoside phosphorylase deficiency caused by a single base change. *J. Biol. Chem.* **262**, 2332–2338 (1987).
 191. Furman, R. R. & Hoelzer, D. Purine Nucleoside Phosphorylase Inhibition as a Novel Therapeutic Approach for B-Cell Lymphoid Malignancies. *Semin. Oncol.* **34**, S29–S34 (2007).
 192. Schramm, V. L. Development of transition state analogues of purine nucleoside phosphorylase as anti-T-cell agents. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1587**, 107–117 (2002).
 193. Ho, M.-C. *et al.* Four generations of transition-state analogues for human purine nucleoside phosphorylase. *Proc. Natl. Acad. Sci.* **107**, 4805–4812 (2010).
 194. Decherchi, S., Berteotti, A., Bottegoni, G., Rocchia, W. & Cavalli, A. The ligand binding mechanism to purine nucleoside phosphorylase elucidated via molecular dynamics and machine learning. *Nat. Commun.* **6**, 6155 (2015).
 195. Miles, R. W., Tyler, P. C., Furneaux, R. H., Bagdassarian, C. K. & Schramm, V. L. One-Third-the-Sites Transition-State Inhibitors for Purine Nucleoside Phosphorylase †. *Biochemistry* **37**, 8615–8621 (1998).
 196. Lebon, G. *et al.* Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* **474**, 521–525 (2011).
 197. Jaakola, V.-P. *et al.* The 2.6 Ångstrom Crystal Structure of a Human A_{2A} Adenosine Receptor Bound to an Antagonist. *Science (80-.)*. **322**, 1211–1217 (2008).
 198. Gutiérrez-de-Terán, H. *et al.* The Role of a Sodium Ion Binding Site in the Allosteric Modulation of the A2A Adenosine G Protein-Coupled Receptor. *Structure* **21**, 2175–2185 (2013).
 199. Gao, Z.-G. & Ijzerman, A. P. Allosteric modulation of A2A adenosine receptors by amiloride analogues and sodium ions. *Biochem. Pharmacol.* **60**, 669–676 (2000).
 200. Katritch, V. *et al.* Allosteric sodium in class A GPCR signaling. *Trends Biochem. Sci.* **39**, 233–244 (2014).
 201. Selent, J., Sanz, F., Pastor, M., De Fabritiis, G. & Fabritiis, G. De. Induced Effects of Sodium Ions on Dopaminergic G-Protein Coupled Receptors. *PLoS Comput. Biol.* **6**, e1000884 (2010).
 202. Piirainen, H., Ashok, Y., Nanekar, R. T. & Jaakola, V.-P. Structural features of adenosine receptors: From crystal to function. *Biochim. Biophys. Acta - Biomembr.* **1808**, 1233–1244 (2011).
 203. Fredholm, B. B., Ijzerman, A. P., Jacobson, K. A., Linden, J. & Müller, C. E. International Union of Basic and Clinical Pharmacology. LXXXI. Nomenclature and Classification of Adenosine Receptors—An Update. *Pharmacol. Rev.* **63**, 1–34 (2011).
 204. Schwartz, T. W., Frimurer, T. M., Holst, B., Rosenkilde, M. M. & Elling, C. E. Molecular mechanism of 7TM receptor activation—A global toggle switch model. *Annu. Rev. Pharmacol. Toxicol.* **46**, 481–519 (2006).
 205. Shi, L. & Javitch, J. A. The binding site of aminergic G protein-coupled receptors: the transmembrane segments and second extracellular loop. *Annu. Rev. Pharmacol. Toxicol.* **42**, 437–67 (2002).
 206. Ohren, J. F. *et al.* Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe

- novel noncompetitive kinase inhibition. *Nat. Struct. Mol. Biol.* **11**, 1192–1197 (2004).
207. La Sala, G. *et al.* HRD Motif as the Central Hub of the Signaling Network for Activation Loop Autophosphorylation in Abl Kinase. *J. Chem. Theory Comput.* **12**, 5563–5574 (2016).
 208. Iacob, R. E. *et al.* Conformational disturbance in Abl kinase upon mutation and deregulation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1386–1391 (2009).
 209. Iacob, R. E., Zhang, J., Gray, N. S. & Engen, J. R. Allosteric Interactions between the Myristate- and ATP-Site of the Abl Kinase. *PLoS One* **6**, e15929 (2011).
 210. Dixit, A. & Verkhivker, G. M. Computational modeling of allosteric communication reveals organizing principles of mutation-induced signaling in ABL and EGFR kinases. *PLoS Comput Biol* **7**, e1002179 (2011).
 211. Suarez, J. *et al.* Catalytic site conformations in human PNP by 19F-NMR and crystallography. *Chem. Biol.* **20**, 212–222 (2013).
 212. Reynolds, K. A., McLaughlin, R. N. & Ranganathan, R. Hot Spots for Allosteric Regulation on Protein Surfaces. *Cell* **147**, 1564–1575 (2011).
 213. Quek, L. S., Bolen, J. & Watson, S. P. A role for Bruton's tyrosine kinase (Btk) in platelet activation by collagen. *Curr Biol* **8**, 1137–1140 (1998).
 214. Honda, F. *et al.* The kinase Btk negatively regulates the production of reactive oxygen species and stimulation-induced apoptosis in human neutrophils. *Nat. Immunol.* **13**, 369–378 (2012).
 215. Liu, X. *et al.* Intracellular MHC class II molecules promote TLR-triggered innate immune responses by maintaining activation of the kinase Btk. *Nat. Immunol.* **12**, 416–424 (2011).
 216. Satterthwaite, A. B. & Witte, O. N. The role of Bruton's tyrosine kinase in B-cell development and function: a genetic perspective. *Immunol. Rev.* **175**, 120–127 (2000).
 217. Tsukada, S. *et al.* Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. *Cell* **72**, 279–290 (1993).
 218. Burger, J. A. Bruton's tyrosine kinase (BTK) inhibitors in clinical trials. *Current Hematologic Malignancy Reports* **9**, 44–49 (2014).
 219. Hendriks, R. W., Yuvaraj, S. & Kil, L. P. Targeting Bruton's tyrosine kinase in B cell malignancies. *Nat. Rev. Cancer* **14**, 219–232 (2014).
 220. Bradshaw, J. M. The Src, Syk, and Tec family kinases: Distinct types of molecular switches. *Cellular Signalling* **22**, 1175–1184 (2010).
 221. Hyvonen, M. & Saraste, M. Structure of the PH domain and Btk motif from Bruton's tyrosine kinase: molecular explanations for X-linked agammaglobulinaemia. *EMBO J.* **16**, 3396–3404 (1997).
 222. Wang, Q. *et al.* Autoinhibition of Bruton's tyrosine kinase (Btk) and activation by soluble inositol hexakisphosphate. *Elife* **4**, 96–100 (2015).
 223. Marquez, J. A. *et al.* Conformation of full-length Bruton tyrosine kinase (Btk) from synchrotron X-ray solution scattering. *EMBO J.* **22**, 4616–4624 (2003).
 224. Park, H. *et al.* Regulation of Btk function by a major autophosphorylation site within the SH3 domain. *Immunity* **4**, 515–525 (1996).
 225. Rawlings, D. J. *et al.* Activation of BTK by a phosphorylation mechanism initiated by SRC family kinases. *Science (80-.)*. **271**, 822–825 (1996).
 226. Tzeng, S.-R. *et al.* Stability and peptide binding specificity of Btk SH2 domain: Molecular basis for X-linked agammaglobulinemia. *Protein Sci.* **9**, 2377–2385 (2000).
 227. Joseph, R. E., Severin, A., Min, L., Fulton, D. B. & Andreotti, A. H. SH2-Dependent Autophosphorylation within the Tec Family Kinase Itk. *J. Mol. Biol.* **391**, 164–177 (2009).
 228. Round, A. R. *et al.* Automated sample-changing robot for solution scattering experiments at the EMBL Hamburg SAXS station X33. *J. Appl. Crystallogr.* **41**, 913–917 (2008).
 229. Franke, D. & Svergun, D. I. DAMMIF , a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).
 230. Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).
 231. Nguyen, H., Maier, J., Huang, H., Perrone, V. & Simmerling, C. Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *J. Am. Chem. Soc.* **136**, 13959–13962 (2014).
 232. de Vries, S. J., van Dijk, A. D. J. & Bonvin, A. M. J. J. WHISCY: What information does surface conservation yield? Application to data-driven docking. *Proteins Struct. Funct. Bioinforma.* **63**, 479–489 (2006).
 233. Svergun, D., Barberato, C., Koch, M. H. J. & IUCr. CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).

234. Wriggers, W. Conventions and workflows for using Situs. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 344–351 (2012).
235. Kjoller, L. & Hall, A. Signaling to Rho GTPases. *Exp. Cell Res.* **253**, 166–179 (1999).
236. Vignal, E. *et al.* Characterization of TCL, a new GTPase of the Rho family related to TC10 and Cdc42. *J. Biol. Chem.* **275**, 36457–36464 (2000).
237. de Toledo, M. *et al.* The GTP/GDP cycling of rho GTPase TCL is an essential regulator of the early endocytic pathway. *Mol. Biol. Cell* **14**, 4846–56 (2003).
238. Ho, H. *et al.* RhoJ regulates melanoma chemoresistance by suppressing pathways that sense DNA damage. *Cancer Res.* **72**, 5516–5528 (2012).
239. Ho, H. *et al.* RhoJ modulates melanoma invasion by altering actin cytoskeletal dynamics. *Pigment Cell Melanoma Res.* **26**, 218–225 (2013).
240. Kartal Yandim, M., Apohan, E. & Baran, Y. Therapeutic potential of targeting ceramide/glucosylceramide pathway in cancer. *Cancer Chemotherapy and Pharmacology* **71**, 13–20 (2013).
241. Ogretmen, B. & Hannun, Y. A. Biologically active sphingolipids in cancer pathogenesis and treatment. *Nat. Rev. Cancer* **4**, 604–616 (2004).
242. Liu, Z. *et al.* Tandem ¹⁸O Stable Isotope Labeling for Quantification of N-Glycoproteome. *J. Proteome Res.* **9**, 227–236 (2010).
243. Mahdy, A. E. M. *et al.* Acid ceramidase upregulation in prostate cancer cells confers resistance to radiation: AC inhibition, a potential radiosensitizer. *Mol. Ther.* **17**, 430–8 (2009).
244. Seelan, R. S. *et al.* Human acid ceramidase is overexpressed but not mutated in prostate cancer. *Genes Chromosom. Cancer* **29**, 137–146 (2000).
245. Canals, D., Perry, D. M., Jenkins, R. W. & Hannun, Y. A. Drug targeting of sphingolipid metabolism: Sphingomyelinases and ceramidases. *British Journal of Pharmacology* **163**, 694–712 (2011).
246. Morad, S. A. F. & Cabot, M. C. Ceramide-orchestrated signalling in cancer cells. *Nat. Rev. Cancer* **13**, 51–65 (2012).
247. Realini, N. *et al.* Discovery of highly potent acid ceramidase inhibitors with in vitro tumor chemosensitizing activity. *Sci. Rep.* **3**, 1035 (2013).
248. Draper, J. M. *et al.* Discovery and Evaluation of Inhibitors of Human Ceramidase. *Mol. Cancer Ther.* **10**, 2052–2061 (2011).
249. Pizzirani, D. *et al.* Benzoxazolone carboxamides: Potent and systemically active inhibitors of intracellular acid ceramidase. *Angew. Chemie - Int. Ed.* **54**, 485–489 (2015).

Publications

- 1) La Sala, G., Riccardi, L., Gaspari, R., Cavalli, A., Hantschel, O., De Vivo, M. HRD Motif as the Central Hub of the Signaling Network for Activation Loop Autophosphorylation in Abl Kinase. *J. Chem. Theory Comput.* **12**, 5563–5574 (2016). (Cover article)
- 2) La Sala, G., Decherchi, S., De Vivo, M., Rocchia, W. Automated analysis of binding pocket dynamics and allosteric networks in proteins. (In preparation)
- 3) La Sala, G., Duarte, D., Lamontanara, A. J., Hantschel, O., De Vivo, M., Structural characterization of BTK kinase via integrative modeling techniques. (In preparation)
- 4) Ortega, J., Arencibia, J., La Sala G., Borgogno, M., Bono, L., Armirotti, M., Giroto, S., Ganesan, A., De Vivo M., Pharmacophore design and scaffold exploration to discover novel, potent, and chemically stable inhibitors of acid ceramidase in melanoma cells. (Submitted)

Communications to conferences

- 1) *Biophysical Society 61st Annual Meeting*. New Orleans (LA), February 11-15 2017. “Molecular simulations to unravel the allosteric interplay between the SH2 domain and A-loop plasticity in protein kinases” [Talk]
- 2) *21st EuroQsar*. Verona, September 4-8 2016. “Allosteric interplay between the SH2 domain and the DFG/HRD motifs modulates A-loop phosphorylation in Abl kinase” [Poster]
- 3) *Models for protein dynamics 1976-2016*. Lausanne, February 15-18 2016. “NanoShaper pocket Tracker: a new tool to identify and analyze pockets along MD simulations” [Poster]
- 4) *Chemical and genomics-based strategies in the discovery of novel drug targets*. University of Bologna, June 22-26 2015. “Understanding protein kinases regulation through computational methods” [Talk]
- 5) *Computational advances in Drug Discovery*. Lausanne, September 22-25 2015. “Detection and evaluation of allosteric pockets and their networks from MD simulations” [Talk]
- 6) *4th Meeting Computational Driven Drug Discovery*. Pomezia, February 24-26 2015. “Insight into the allosteric inhibition of and regulation of Abl kinase” [Poster]