Exploring the role of conformational dynamics in the regulation of tyrosine kinases.

Silvia Lovera

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Chemistry

University College London

16 November 2015

I, Silvia Lovera, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

A mamma e nonna

Abstract

Tyrosine kinases (TKs) are a family of signalling proteins of great pharmaceutical importance, as they are involved in the regulation of most cellular pathways. TKs catalytic activity is strictly regulated by conformational changes and post-translational modifications, and their deregulation is involved in numerous human diseases, ranging from cancer to autoimmune diseases. Among tyrosine kinases, Abl and Src are of particular interest for cancer research. The Abl domain in the BCR-Abl fusion protein is the main cause of chronic myeloid leukemia, and it was the target of the first successful anti-leukemic therapy, the powerful kinase inhibitor imatinib. We now know that imatinib effectively inhibits BCR-Abl, as well as Kit and Lck kinases, by binding to a specific inactive state, in which the conserved Asp-Phe-Gly motif (DFG) assumes a peculiar "out" conformation. Still, there are many questions on its mode of action. For instance, other TKs with an extended identity with Abl (such as Src, which has 45% sequence identity) bind much less strongly to imatinib, in spite of very similar binding mode. Moreover, the mode of action of drug-resistant mutations that induce imatinib resistance and cause an increasing number of relapses in patients under treatment, is still poorly understood. Understanding the molecular mechanisms responsible for the observed differences in imatinib activity, is essential for the development of new selective anticancer drugs. In this thesis, by using computational and experimental approaches, I have investigated the reasons leading to drug resistance and the differential binding affinity in homologous TKs. A combination of enhanced sampling molecular dynamics simulations (such as parallel tempering metadynamics or PTmetaD) were used to reconstruct and compare the free energy landscape associated with the relevant conformational changes. Mutagenesis and isothermal titration calorimetry were used to validate the computational results.

The work presented in this thesis has led to the following publications:

1. Silvia Lovera, Ludovico Sutto, Ralitza Boubeva, Leonardo Scapozza, Nicole Dölker and Francesco L. Gervasio, *The different flexibility of c-Src and c-Abl kinases regulates the accessibility of a druggable inactive conformation*, Journal of the American Chemical Society 134, 2496-2499 (2012). DOI:10.1021/ja210751t

2. Silvia Lovera, Maria Morando, Encarna Pucheta-Martinez, Jorge L. Martinez-Torrecuadrada, Giorgio Saladino and Francesco L. Gervasio, *Towards a molecular understanding of the link between imatinib resistance and kinase conformational dynamics*, Plos Computational Biology 11(11), e1004578 (2015). DOI:10.1371/journal.pcbi.1004578

3. Maria Morando, Giorgio Saladino, Nicola D'Amelio, Silvia Lovera, Moreno Lelli, Marco Marenchino, Blanca López, Ramón Campos-Olivas and Francesco L. Gervasio, *Conformational dynamics and unfolding associated with the binding of an anticancer drug to the c-Src kinase*, in preparation.

Acknowledgements

First of all I would like to thank my supervisor Prof. Francesco Luigi Gervasio not only because he gave me the opportunity to join his group and doing my PhD, but also for his guidance and support throughout this project and in life. His guide and passion for science inspired me in my research and now in my career.

I remember the first day of PhD, it was also the first time I saw a protein kinase. Since that very first day, my fascination for these biological systems has grown bigger and bigger over the years, together with my love for science, and this is another of the many things for which I have to be grateful to my supervisor.

My PhD has been not always easy, but every challenge I have overcome has made me a stronger scientist and a stronger person. There are no words to express how thankful I am for all the experiences I lived and the people I have met along this incredible journey.

During these years, I have met and work side by side with so many amazing professionals from whom I have learnt and discuss of science. A particular thank goes to my closest collaborators and mentors, Dr Nicole Dölker, Dr Giorgio Saladino, Dr Maria Morando, Dr Ludovico Sutto and Encarna Pucheta-Martinez, for their substantial contribution at various points of the work presented in this thesis, and not less important for their support and help all the way through. I was also fortunate to collaborate with many other people over the course of my PhD, Dr Jorge Martinez-Torrecuadrada, Dr Marco Marenchino, Prof Leonardo Scapozza, Dr Remo Perozzo and Ralitza Boubeva, masters in protein expression and purification.

I would like to thank all present and past group members. I shared my days not only with some of the most passionate and inspiring scientists I have never known, but also with amazing people and friends for life. I remember with great pleasure my time in Madrid with Marta, Simone, Antonio, Nicola, Ilaria, and in London with Kristen and

Vladas.

Throughout these years lived in between these two cities, I was lucky to meet many special people now dear friends Francesca, Deborah, Alessandra, Chiara, Enrico, Ugo, Ara, Marija, Cristina and Francesco. To all and each one of them I say a big thank you, for bringing me the pleasure of enjoying life. How to forget my amazing life-long friends Elena, Elisa, Marta and Valentina, who have been ever-present, no matter how far we were, they never made me feel alone.

Lastly, a special thank goes to Juan, whose love and support have been of inestimable importance, and to my family papá, nonno, Sara, Marco and little Aileen. Thank you for your unconditional love, for being my anchor and my inspiration, but most importantly because you taught me to love, to be good and to follow my dreams.

Contents

Preface **25**

1 Introduction **26**

1.1 The modern view of dynamics and allostery in proteins 26

1.1.1 Dynamics of proteins: the link between structure and function . 26

1.1.2 The hierarchy of dynamics in proteins: motions on different timescales 27

1.1.3 The integrated view of protein allostery 28

1.2 Tyrosine kinases as paradigms of highly dynamic structures 29

1.2.1 Structure and conformational changes in tyrosine kinases 30

1.2.2 Allostery in tyrosine kinases 32

1.3 Tyrosine kinases and their deregulation in cancer 34

1.3.1 The tyrosine kinase family 34

1.3.2 The Src tyrosine kinase 35

1.3.3 The Abl tyrosine kinase 36

1.3.4 Other tyrosine kinases of interest 37

1.4 Tyrosine kinases as anti-cancer drugs targets 38

1.4.1 Chromic myeloid leukemia and imatinib "promiscuity" 38

1.4.2 Binding of Imatinib: conformational selection or induced fit? . . 41

1.5 Drug-resistance in tyrosine kinases 45

1.5.1 Imatinib and resistance 45

1.5.2 Alternative strategies and drugs to overcome resistance 48

2 Materials and Methods **51**

2.1 Computational Methods 52

2.1.1 Molecular Dynamics . . . 52
2.1.2 Enhanced sampling algorithms and free energy calculations . . 54
2.1.3 Metadynamics and Parallel-Tempering Metadynamics 55
2.1.4 Classical MD simulations details . . . 58
2.1.5 Trajectory analysis . . . 59
2.1.6 Enhanced sampling simulations details . . . 62
2.1.7 Rational choice of PTmetaD parameters . . . 66
2.1.8 PTmetaD analysis . . . 69
2.1.9 Limitations of the computational methods . . . 71
2.2 Experimental Methods . . . 72
2.2.1 *In vitro* protein expression . . . 72
2.2.2 Protein purification . . . 73
2.2.3 Expression and purification of Src wild-type and Src engineered mutants . . . 76
2.2.4 Isothermal titration calorimetry . . . 79

3 The different flexibility of Src and Abl kinases regulates the accessibility of the DFG-out conformation 83
3.1 Classical MD simulations and RMSF analysis of Src and Abl . . . 84
3.2 PTmetaD calculation and free energy surface reconstruction . . . 86
3.3 Free energy reweighting . . . 91
3.4 Solvent accessible surface area and water accessibility . . . 93
3.5 Calculation of the pK_a and influence of Asp404 protonation . . . 94
3.6 Discussion . . . 94

4 The correlation of sub-μs dynamics to drug binding in tyrosine kinases 96
4.1 Selection of the tyrosine kinases included in the study . . . 97
4.2 Allosteric coupling analysis . . . 97
4.3 Classical MD simulations and RMSF analysis of the homologous TKs . 99
4.4 Principal component analysis . . . 102
4.5 Experimental validation . . . 104
4.6 Discussion . . . 106

5 Src mutagenesis confirms the link between flexibility and imatinib binding affinity 108

5.1 Selection of the Src mutants included in the study 109

5.2 MD simulations and RMSF analysis of the engineered mutants of Src . 109

5.3 Expression and purification of the engineered mutants of Src 116

5.4 ITC experiments of the engineered mutants of Src 116

5.5 Discussion 120

6 Understanding of the molecular mechanism of imatinib resistance 123

6.1 Selection of the Abl resistant mutants included in the study 124

6.2 Classical MD simulations and RMSF analysis of the resistant mutants of Abl 125

6.3 Principal component analysis 127

6.4 PTmetaD-WTE calculation and free-energy surface reconstruction . . . 128

6.5 Estimating the entropic and enthalpic contribution to the ΔG_{in-out} . . . 134

6.6 Free energy reweighting 136

6.7 Binding free-energy of imatinib to Abl and T315I mutant 138

6.8 Discussion 140

7 Conclusions 143

Bibliography 146

List of Figures

1.1 Free energy profile following the different timescales and amplitude of motions in proteins, as in the classification made by Frauenfelder *et al.*[2, 3]. Slow conformational changes happen at higher free energy values (rare events), at slower timescale (tier-0) and they are characterised by motions of bigger amplitude. Faster motions could be found in tier-1 (ns timescale) and in tier-2 (fs), mostly associated with bonds vibration and side-chains rotation. Figure taken from Ref [2]. 28

1.2 Representations of the reshaping that the energy landscapes of proteins could experience after ligand binding. In black is the energy profile before ligand binding, in red the new profile upon ligand binding. (a) The ligand selects for a conformation among many equally stable conformations. (b) The ligand binds to a stable native conformation of a protein determining a conformational change that leads to another stable conformation. (c) The ligand binds to a stable native conformation bringing to a more heterogeneous landscape of accessible conformations. Figure taken from Ref [10]. 29

1.3 Front view of the Abl kinase domain (a) with highlighted in different colours the most important structural features and in sticks the tyrosine residue of the activation loop (in green), which phosphorylation activates the kinase. (b) Side view of Abl with a more detailed view of the DFG motif (in orange sticks) and of its two possible conformations, DFG-in and DFG-out. 30

1.4 Schematic representation of the possible conformations a tyrosine kinase can assume. There is just one active conformation, with well defined structural characteristics to fulfil (DFG-in, A-loop open, αC-helix Glu-in) but many inactive intermediate conformations, with different arrangements the catalytic important elements can adopt. Among them just one is of interest in the context of this thesis work, and is the inactive DFG-out conformation in which the drug imatinib (in magenta sticks) binds to the kinase active site (A-loop closed, αC-helix Glu-in). . 31

1.5 Regions of the kinase domain involved in the allosteric network of TKs. The different colours identify every element of the network: P-loop in pink, αC-helix in cyan, DFG motif in orange, A-loop in green and the αG-helix in wheat. Imatinib is shown in red sticks binding to the kinase active site. 32

1.6 Protein Kinase A (PKA) structure where hydrophobic spines have been detected for the first time. View of kinase domain with the catalytic spine (C-spine) shown in yellow and the regulatory spine (R-spine) shown in red. Figure taken from Ref [23]. 33

1.7 Catalytic cycle of a tyrosine kinase protein. Figure taken from Ref [26]. 34

1.8 Structure of Src kinase. The full-length kinase is constituted by three domains: SH2, SH3 and kinase domain (KD, the only considered in this study). The different spatial orientation of the domains discriminates between the inactive and active forms, adoptable by Src. For the activation or inactivation of the full-length kinase, crucial are two sites of phosphorylation: Tyr527 of the C-terminal tail (phosphorylated in the inactive form) and Tyr416 of the activation loop (phosphorylated in the active form). Figure taken from Ref [34]. 35

1.9 Cartoon representation of the Abl structure. (a) The full-length kinase is constituted by three domains: SH2, SH3 and kinase domain (KD, the only considered in the study). The kinase is in the inactive form with the myristate tail (Myr) docked to the hydrophobic pocket in the C-lobe. (b) Cartoon representation of the activated form of Abl, with the peculiar position adopted by the SH2 domain, on top of the kinase domain. Figure taken from Ref [41]. 36

1.10 Binding mode of imatinib to Src and Abl obtained by x-ray crystallography. In sticks is shown the DFG motif adopting the "out" conformation. Imatinib adopts the same binding pose in both kinases. Figure taken from Ref [20]. 39

1.11 X-ray structure of Abl in the DFG-out conformation with imatinib bound. At a side, detailed image of the P-loop region. To be noted is the kinked conformation adopted by the P-loop (in red) and ascribed to reorientation of residue Tyr253 (in sticks). In Src and Kit instead of a tyrosine there is a phenylalanine. Coloured in green the αC-helix and in yellow the A-loop. Imatinib is shown in grey sticks. Figure taken from Ref [20]. 40

1.12 (A) pH dependence of imatinib binding to Abl wild-type (in black). (B) Variation of the percentage of the DFG-out population of Abl corresponding to different pH values (in black). (C) Binding rates of imatinib to Abl WT at different pH show high pH dependence for the binding. (D) Binding rates of dasatinib to Abl WT at different pH show weak pH dependence (dasatinib binds to the DFG-in conformation). Figure taken from Ref [70]. 41

1.13 Structure of the drug imatinib. In dotted lines are shown the interactions it makes when bound to the kinase. The residues are common to Src and Abl. Figure taken from Ref [72]. 42

1.14 Imatinib binding to the active site of the kinase occupying both the ATP pocket and the allosteric pocket. When imatinib is bound, the kinase is adopting the DFG-out, inactive, conformation. Coloured in green is the DFG motif, in yellow the A-loop. 42

1.15 Free-energy surfaces of Src KD with A-loop not phosphorylated (a) and phosphorylated (b). The not phosphorylated surface (a) shows more energy minima while, the phosphorylated form (b) only one, corresponding to the open A-loop conformation. The method used in the study is the Umbrella Sampling enhanced method[74]. On the x-axis, the distance describing the A-loop closed (lower values) to open (higher values). On the y-axis, the subtraction of two distances describing the switch of the αC-helix from the Glu-in (higher values) to Glu-out (lower values) conformation. Figure taken from Ref [73]. 43

1.16 Response curves of drugs binding to the phosphorylated (p) and not phosphorylated (np) forms of Src (a) Abl (b) and Kit (c). Drug affinity has been tested with different methods in the three studies thus, values could not be directly compared. (a) Fraction bound of imatinib to Src "pY412" and "npY412". (b) Response curves of imatinib and dasatinib binding to Abl "p" and "np". It was also tested a mutant of Abl, in which phosphorylation has been impaired by mutating the tyrosine of the A-loop (Y393F). (c) Response curves of imatinib and sunitinib binding to Kit. The label "activated KIT" stands for phosphorylated while, "inactivate KIT" for not phosphorylated. Figure (a) taken from Ref [75], Figure (b) taken from Ref [76] and Figure (c) taken from Ref [77]. 44

1.17 Abl KD (a) with resistant mutations highlighted (b). Green dots are those mutants with a known mechanism of action (*e.g.* elision of crucial bonds for the stability of the binding), while the red dots are the ones whose mechanism is still unknown[93]. 46

1.18 (a) Survival curves of patients expressing non P-loop and P-loop mutations. (b) Abl second messenger STAT5 phosphorylation levels for Abl WT and resistant mutants of Abl (T315I, G250E, G250A). T315I ("gatekeeper" mutant) and G250E (P-loop mutant) are not inhibited at high imatinb doses (above 5 μM, high degree of resistance). Figure (a) taken from Ref [94] and Figure (b) taken from Ref [95]. 47

1.19 Free-energy surfaces of EGFR WT and of three EGFR oncogenic mutants (L858R, T790M and T790M-L858R). The mutations are responsible for a shift in the conformational space sampled by the kinase, leading to the stabilisation of the active conformation (open active-like A-loop and αC-helix Glu-in). Figure taken from Ref [100] 48

1.20 Binding modes of the two major classes of kinase inhibitors, Type I and Type II. Type I drugs bind to the ATP pocket, below the P-loop region, type II drugs occupy both the ATP and the allosteric pockets, adopting a bridge position over the DFG motif at the N-terminal of the A-loop. . 49

2.1 Gaussians are added along the CV space to help the system overcoming energy barriers during the metaD simulation (a). When the entire conformational space is filled with Gaussians (b), the added bias is summed (c) and the FE profile reconstructed (d). Figure taken from Ref [164]. 56

2.2 Schematic example of the exchanges happening among replicas at different temperatures during a Parallel Tempering simulation. Figure taken from Ref [166]. 58

2.3 Graphical representation of the coarse-grained atom displacement as described in the method of Balabin *et al.*[178]. Figure taken from Ref [178]. 62

2.4 Collective variables chosen for the well-tempered PTmetaD simulations of Src and Abl. 2SRC.pdb numbering considered, with the exception of the residue I293 in Abl that corresponds to L317 in Src. In detail, the CV3 dihedral combination of the DFG motif. Figure taken from Ref [93]. 64

2.5 Comparison of the potential energy fluctuations for two replicas in a normal NVT simulation and in a WTE simulation. Figure taken from Ref [184]. 65

2.6 Example of the PLUMED input file (version 1.3[163]) used to perform the PTmetaD simulations described in this thesis. Highlighted in yellow are the parameters chosen by the user. The numbering used for the residues defining the CVs do not correspond to the 2SRC.pdb numbering, used in Section 2.1.6, but to the original pdb file used in the simulation. 68

2.7 Collective variables chosen for the reweighting of the well-tempered PTmetaD simulation for both Src and Abl (2SRC.pdb numbering considered). Figure taken from Ref [187]. 70

2.8 Representation of the activation loop in the open (in red) and closed (in green) conformations used as references in the PTmetaD reweighting of Src and Abl. 71

2.9 Improvement of force-fields accuracy over time. The force-fields with lower scores are those showing a higher agreement with experiments. Figure taken from Ref [132]. 72

2.10 Example of the chromatogram generated by affinity and ion-exchange chromatography method. During purification the solution, before being collected, passes through an UV detector. Thus, the UV absorbance profile is monitored to check for protein concentration. After loading the sample into the column (load phase, no UV signal), the protein of interest will bind to the column, while all the impurities will be eluted, determining the increase of the UV signal. It follows a wash phase to eliminate all impurities, and only when the UV profile has returned to baseline, the conditions will be modified to determine the elution of the target protein (elution phase, UV peak). Figure taken from Ref [199]. . 74

2.11 Ion-exchange chromatography. The matrix is negatively charged (grey beads). Positively charged proteins (blue beads) interacts with the matrix and are retained, while negatively charged proteins (pink beads) are eluted. Figure taken from Ref [201]. 75

2.12 Size-exclusion chromatography. Large proteins are excluded by the column beads, while small proteins are retained and eluted later. Figure taken from Ref [202]. 76

2.13 Schematic workflow of the protocols used for the expression and purification of Src WT and the engineered mutants of Src. In red are highlighted the changes made to the original protocol of Ref [194]. . . . 78

2.14 Schematic representation of an ITC instrument. Figure taken from Ref [207]. 80

2.15 Example of the outputs produced during an ITC experiment. Figure taken from Ref [208]. 81

3.1 (A) RMSF profiles of the C_α atoms of Src (in green) and Abl (in red). (B) RMSF analysis representation directly on the structures of Abl and Src. The thickness of the ribbon is proportional to the RMSF magnitude. The colour spectrum ranges from more rigid (blue) to more flexible (red). Figure taken from Ref [187]. 85

3.2 Conformations adopted by the αC-helix of Src (in green) and Abl (in red). The distances F405-L317 in Src and F405-I293 in Abl (residues shown in sticks) were monitored to characterise the αC-helix movement. 85

3.3 (A) Distances Phe-Leu of Src (in green) and Phe-Ile of Abl (in red) were monitored over the 1 μs long simulation. (B) Distribution of Src and Abl populations as function of the residue pair distance. Figure taken from Ref [187]. 87

3.4 1-D projection over time of the free-energy basins of CV1 (A) and CV2 (B) of Abl and Src (C, D). Figure taken from Ref [187]. 88

3.5 (A) Reconstructed FESs of Src (left) and Abl (right). The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. (B) Structures corresponding to the relevant minima and transition states. Figure taken from Ref [187]. 89

3.6 (A) Free energy profiles projected along CV2, distance Phe-Leu in green for Src and Phe-Ile in red for Abl. The green and red bars indicate the width of the respective minimum (B) Conformations adopted by the αC-helix of Src (in green) and Abl (in red), corresponding to the local energy minimum indicated with a green and red star, respectively, in panel A. 90

3.7 Alignment of structures extracted from the PTmetaD minima (DFG-in and DFG-out). Changes in the N-lobe loops and αC-helix conformations of Abl (in red) can be appreciated in the passage from DFG-in to DFG-out while, slight differences can be detected for Src (in green). . . 91

3.8 (A) Reweighting CVs (only Src numbering used). (B) 1-D projections of the reweighted FE profiles for Src (in green) and Abl (in red). The green and red bars underline the minima of Src and Abl, respectively. The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. Figure taken from Ref [187]. 92

3.9 Comparison of the cavity dimensions in the DFG-in and DFG-out conformations of Src (in green) and Abl (in red). 93

3.10 Water molecules within 5 Å of the DFG motif of Abl (shown in sticks) in the DFG-in and DFG-out conformation. 94

4.1 Allosteric coupling maps of Src, Abl and the homologous TKs together with the structure of the kinase domain with highlighted the elements involved in the allosteric coupling. In the maps, the A-loop region corresponds to the area included in the white dashed lines. Figure taken from Ref [93]. 98

4.2 (a) RMSF profiles of Src (blue) and Abl (red). Grey panels identify the N-lobe and the A-loop region. (b) RMSF profiles of the homologous TKs. In the two panels are reported the RMSF profiles of the N-lobe region (left) and A-loop region (right), being the areas whose flexibility varies most among the pool of TKs. Shades of red identify the flexible TKs, shades of blue the more rigid ones. Dotted lines are used for clarity. Figure taken from Ref [93]. 100

4.3 RMSF profile of the αG-helix of Src, Abl and the homologous TKs. Shades of red identify the flexible TKs, shades of blue the more rigid ones. Dotted lines are used for clarity. Figure taken from Ref [93]. . . . 101

4.4 Schematic representation of the KD dominant motions: the hinge-bending motion (a) and the lobe twist (b). Both are characterised by movements of the N-lobe relative to the C-lobe. 102

4.5 PCA analysis. Comparison of the conformational space explored by the homologous TKs under study (Src, Kit, Lck, Met and EGFR) with respect to Abl. The first eigenvector (Abl eig 1) describes the hinge motion, while the second eigenvector (Abl eig 2) the lobe twist. 103

4.6 The activity matrix shows the residual catalytic activity for each kinase-drug match retrieved from the results obtained in the work of Ref. [213]. Shades of green identify the kinases effectively inhibited by the considered drug, so having a low residual catalytic activity, while shades of red identify kinases not inhibited by the drug and maintaining an high residual activity. With residual activity is meant the ability of the kinase to phosphorylate biological substrates. The drugs tested in the study have been grouped according to their binding mode to the kinase (type I, type II). To the other groups belong the EGFR specific inhibitors and the nonspecific inhibitors). 106

5.1 Structure of the kinase KD with highlighted the positions of the engineered mutants of Src. The mutations affecting Src dynamics are coloured in orange, the ones affecting Src contacts are coloured in cyan. 110

5.2 RMSF of the engineered single mutants of Src (Q275A, P299Q, Q420E, V461I). The RMSF profile of every mutant (dashed orange line) is plotted together with the two references profiles of Lck (green line, flexible) and Src WT (blue line, rigid). The orange dots identify the site of the mutation and the orange bars the regions of the structure whose fluctuation results affected by the mutation. 112

5.3 (a) RMSF profiles of Src (blue) and Abl (red). Grey panels identify the N-lobe and the A-loop region. (b) RMSF profiles of the engineered mutants of Src. In the two panels are reported the RMSF profiles of the N-lobe region (left) and A-loop region (right). Shades of red identify the more flexible mutants, shades of blue the more rigid. Dotted lines are used for clarity. Figure taken from Ref [93]. 113

5.4 RMSF profile of the αG-helix of Src, Abl and the engineered mutants of Src. Shades of red identify the more flexible mutants, shades of blue the more rigid. Dotted lines are used for clarity. Figure taken from Ref [93]. 114

5.5 RMSF of the engineered mutants of Src: M302G and E415G/Y416G/V461S. The RMSF profile of the mutant (cyan dotted line) plotted with the two references profiles of Abl (red line, flexible) and Src WT (blue line, rigid). The dots in cyan identify the sites of the mutation and the bars the regions of the structure whose fluctuation results affected by the mutation. 115

5.6 (a) HisTrap column chromatogram: the protein of interest is the more intense UV signal. (b) size-exclusion S75 column chromatogram: the protein of interest corresponds to the last signal. (c) Gels identifying the obtained mutants of Src after purification (around 35 kDa). Figure taken from Ref [93]. 117

5.7 ITC curves resulting from the experiments performed and reporting the imatinib binding curves towards Src WT and the engineered mutants of Src. Figure taken from Ref [93]. 118

5.8 Figure taken from Ref [20]. ITC curves of imatinib binding to Abl (A) and Src (B). In (C) K_d values of imatinib obtained for Abl WT, Abl phosphorylated and Abl L384G mutant. In (D) K_d values of imatinib binding to Src WT and to some of the Src mutants described in the Ref [20]. Noticeable is the choice of the mutated residues, all in the immediate vicinity of the DFG motif (Asp404-Phe405)[20]. 119

5.9 αG-helix conformations (A, B) sampled in the trajectory of Src WT. Figure taken from Ref [93]. 120

5.10 Structure highlighting the location of the engineered mutants of Src. The mutations affecting Src dynamics are coloured in orange, the ones affecting Src contacts and structure are coloured in cyan. Arrows are illustrating the major allosteric pathways identified by the mutagenesis study, coloured accordingly to mutant. The connection between P-loop and αG-helix is a feature of the Src structure. 121

6.1 (A) IC_{50} values of imatinib corresponding to the resistant mutants of Abl collected in the work of Azam *et al.*[88], compared to the IC_{50} of Abl (in red). The mutations are divided in those lying in flexible regions of the Abl KD (in orange) and those lying in more constrained areas (in green), also directly highlighted on the Abl structure (B). As it could be noticed the mutations on flexible regions (P-loop, A-loop, αC-helix, αG-helix) have a more drastic impact on the IC_{50} (higher is the value, lower is the inhibitory effect of imatinib towards Abl). 124

6.2 (a) RMSF profiles of Src (blue) and Abl (red). Grey panels identify the N-lobe and the A-loop with αG-helix region. (b) RMSF profiles of the resistant mutants of Abl. In the two panels are reported the RMSF profiles of the N-lobe region (left) and A-loop with αG-helix region (right), being the areas whose flexibility varies most among the pool of TKs. Shades of red identify the flexible mutants, shades of blue the more rigid ones. Dotted lines have been used for clarity. Figure taken from Ref [93]. 126

6.3 PCA analysis. Comparison of the conformational space explored by the resistant mutants of Abl with respect to Abl WT. 128

6.4 Reconstructed FESs for Abl WT, Src and the resistant mutants of Abl in function of CV1 and CV2. The DFG-in minimum is identified by the "IN" label, the DFG-out minimum by the "OUT" label and the intermediate Asp-down conformation, when present, by the label "Asp-down". The FESs of Src and Abl WT were taken from Chapter 3 and readapted. The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. Figure taken from Ref [93]. 130

6.5 2D projections showing the convergence of the free-energy of the resistant mutants of Abl (E450K,H396P and E279K). The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. Figure taken from Ref [93]. 131

6.6 2D projections showing the convergence of the free-energy of the resistant mutants of Abl (G250E and T315I). The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. Figure taken from Ref [93]. 132

6.7 DFG out-out conformation identified in the DFG-out minima of G250E (a) and E450K (b). Unfolding of the αC and αG helices in G250E (c), E279K (d), H396P (e, f) and E450K (g). Figure taken from Ref [93]. . . 133

6.8 Structures of Src WT (a), G250E (b), E279K (c), H396P (d), E450K (e) and T315I (f) showing the formation of a helix turn in the A-loop of the resistant mutants of Abl, resembling Src. In sticks is shown the DFG motif and in H396P mutant in red the site of the mutation, located exactly after the helix turn. Figure taken from Ref [93]. 133

6.9 Linear regressions of Abl WT, Src and the resistant mutants of Abl. Each point of the regression corresponds to the ΔG_{in-out} calculated for each replica of the system at different temperature. This was repeated for the PTmetaD simulations performed. Figure taken from Ref [93]. . . 134

6.10 RMSF profiles of the DFG-in (in orange) and DFG-out (in green) conformations for Abl WT, Src and the reisstant mutants of Abl. Figure taken from Ref [93] 135

6.11 Reweighted FESs of Abl WT, Src and the resistant mutants of Abl. Projections along two path CVs describing the opening of the A-loop in Src (CV1) and Abl (CV2). The energy minimum corresponding to open, semi-open and closed A-loop conformations, when present, are labelled with A, B and C, respectively. The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. Figure taken from Ref [93]. 137

6.12 Reconstruction of the (un-)binding FES of imatinib to Abl and T315I. The structures corresponding to the local minima and transition states of the FES are identified with letters. Shown in green stick is the drug imatinib. In the structures of the T315I, the mutation is shown in red sticks. The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. Figure taken from Ref [93]. 139

6.13 Structure with highlighted by red dots the location of the resistant mutants of Abl. Red arrows illustrate the dominant allosteric pathways identified by the analyses performed. The connection between the αC-helix and the A-loop is a feature of the Abl structure. 141

List of Tables

1.1 IC_{50} values of imatinib activity towards each of the targeted TKs. Lower IC_{50} values corresponds to a better imatinib activity (better kinase inhibition). 39

4.1 RMSF of the A-loop region and IC_{50} values of imatinib reported in literature for Abl, Src and the homologous TKs[93]. 102

5.1 RMSF values of the A-loop region, averaged from MD calculations, and the retrieved K_d for Src WT and the engineered mutants of Src by means of ITC (data published in Ref [93]). Lower the K_d, higher the affinity of imatinib for the kinase. The K_d value of Abl was recovered from the work of Seeliger et al.[20]. 118

6.1 Entropic and enthalpic contributions to the ΔG_{in-out} in Abl WT, Src and all the resistant mutants of Abl. The values have been obtained from the linear regressions of the ΔG_{in-out} as a function of temperature shown in Figure 6.9 with equation $\Delta G=\Delta H-T\Delta S$. Figure taken from Ref [93]. 136

Preface

Protein kinases (PKs) are a large family of signalling proteins that regulate a plethora of cellular functions by phosphorylating protein and non-protein substrates. They currently are the most important anti-cancer targets, and we now count more than 100 kinase inhibitors in clinical use or in an advanced development phase [1]. Thanks to the recent technological advancements in drug discovery, including those in computer-aided drug design, more powerful inhibitors have been developed, but the search for selective and effective PKs inhibitors is far from over. This is due, on the one hand, to the complex interplay between conformational changes, protein allostery and drug binding mechanisms. On the other, to the emergence of drug resistance, which makes the most advanced inhibitors and therapies ineffective. To overcome these problems, a deeper comprehension of the regulatory mechanisms in PKs is needed.

In this Ph.D. project, I have studied in great detail the role of the conformational dynamics and cancer-causing mutations in the regulation and de-regulation of selected members of the tyrosine kinase (TK), family of PKs. The TKs studied were chosen for their great historical and biomedical importance. In the first chapter, I summarise the recent literature on the subject. In the second chapter, I describe the computational methods employed and the experimental techniques used to validate the computational results. In the third, fourth and fifth chapters, I discuss the molecular mechanisms responsible for the observed differences in the imatinib activity and in the sixth the mode of action of several drug-resistant mutations. The seventh and last chapter is devoted to final considerations and conclusions.

Chapter 1

Introduction

TKs are very dynamic structures (plastic, flexible), and both their activity and regulation are influenced by coupled movements on different timescales[2, 3]. While fast and localised atomic vibrations are involved in the catalytic activity of TKs, slower conformational changes are crucial to regulation.

In this respect, only recently the role of sub-μs dynamics in proteins has been recognised as crucial in the regulation of conformational changes, binding of ligands and allostery. In this section, I will give a general overview of how the view of proteins as dynamical entities has changed and evolved over the years.

1.1 The modern view of dynamics and allostery in proteins

1.1.1 Dynamics of proteins: the link between structure and function

The view of proteins as dynamic entities has evolved with time. Initially thought as static in the "lock and key" model of enzyme activity[4], they were soon recognised as flexible entities in the "induced fit" theory by Koshland in 1963[5]. Finally in the early 80s, experimental evidence for the dynamics of enzymes both in presence and absence of substrates[6, 7, 8], founded the concept of proteins as flexible systems. For the first time, dynamics was seen as an essential requirement for enzyme activity, and the link between structure and biological function in proteins.

The dynamics of proteins is strictly correlated to the amino acid sequence, with backbone atoms in helices and strands being more restrained than those in loops. Nowadays proteins are seen as in constant fluctuation among different conformations[9, 10]. In particular, conformations adopted by proteins during catalysis seem to be already pre-sampled in the free enzyme, suggesting that the crucial motions for catalysis are intrinsic property of enzymes, deeply encoded in their sequence[11]. Therefore, the biological function is strictly associated with protein dynamics, and the biological processes are mainly determined by alterations in the populations rates of catalytically competent conformations, rather than by on/off mechanisms[3].

Computationally, energy landscapes could be effectively used to describe the conformational space of proteins. Thus the relative probability of every conformation can be defined as well as the energy barrier between one conformation and another. Since protein dynamics is defined by both the timescale and the amplitude of the protein motions, the energy landscapes of proteins turn out to be highly multidimensional[2].

1.1.2 The hierarchy of dynamics in proteins: motions on different timescales

To get a comprehensive description of the energy landscape of proteins, a method that samples their dynamics over a wide range of timescales is required (see Figure 1.1). Kinetically different conformations are usually separated by high energy barriers, thus the interchange among them happen at slow timescales (μs, ms or slower). In Figure 1.1 it can be seen that these slow conformational changes belong to what is defined by Frauenfelder *et al.* as 'tier-0'[2]. The transitions of tier-0 are usually associated to large and rare conformational changes, involved in lots of biological processes. However, within the energy minimum (tier-0), proteins are not static but experience ps to ns fluctuations (tier-1 and tier-2 dynamics)[2]. Proteins can sample a large amount of closely related conformations, separated by low energy barriers. If tier-0 is mostly characterised by large domain motions, tier-1 and tier-2 define fast motions, such as movements of loops and rotation of side-chains.

From experiments (*e.g.* NMR), it has become clear how close is the connection between fast and slow dynamics, since the first is the physical origin of the latter[12, 13, 14, 15]. Changes in the fast dynamics (sub-μs dynamics), closely de-

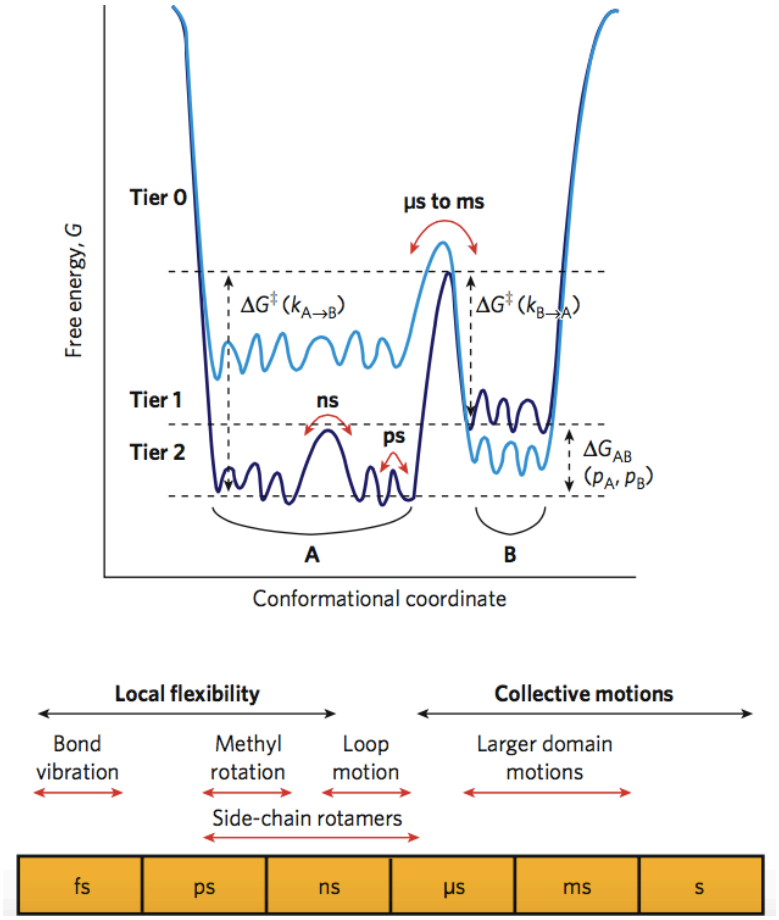

Figure 1.1: Free energy profile following the different timescales and amplitude of motions in proteins, as in the classification made by Frauenfelder *et al.*[2, 3]. Slow conformational changes happen at higher free energy values (rare events), at slower timescale (tier-0) and they are characterised by motions of bigger amplitude. Faster motions could be found in tier-1 (ns timescale) and in tier-2 (fs), mostly associated with bonds vibration and side-chains rotation. Figure taken from Ref [2].

pendent by the amino acid sequence, have an influence on the more complex domain motions of proteins[16]. This hierarchy in the dynamics of proteins, both in time and space (timescale and amplitude of motions), is certainly encoded in the amino acid sequence, but it is also strictly connected to protein allostery[17].

1.1.3 The integrated view of protein allostery

Proteins are able to react to external stimuli via long-range communication, and this communication is essential for every biological event. They interact with tens of substrates, ligands and other proteins at specific sites within their structure. The effect of these interactions is then propagated to regions far away in structure, letting the signal flow. This potent regulatory process, referred as allostery, is also responsible for cru-

cial conformational changes in proteins. A ligand, for example, could provoke changes involved in the ultimate regulation of a biological process, such as the adoption of a functionally active conformation[18].

In the allosteric theory of Monod-Wyman-Changeux in 1965[19], the protein is seen in equilibrium between two major conformations, and the ligand may perturb this equilibrium, binding to one conformation and provoking a conformational change that will stabilise the other (see Figure 1.2, image b). The current view of protein allostery add complexity to this theory, taking into consideration protein dynamics[9]. Today proteins are seen as plastic structures, fluctuating among different conformations. The binding of a ligand determines the reshape of their free energy landscapes, binding to a pre-existing conformation and inducing a shift in the population rates[10] (see Figure 1.2, image a and c).

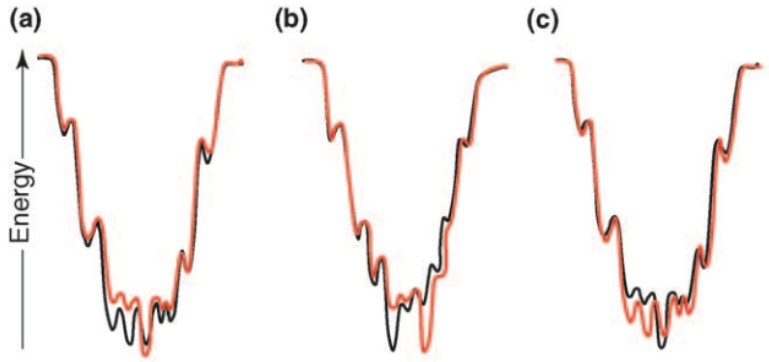

Figure 1.2: Representations of the reshaping that the energy landscapes of proteins could experience after ligand binding. In black is the energy profile before ligand binding, in red the new profile upon ligand binding. (a) The ligand selects for a conformation among many equally stable conformations. (b) The ligand binds to a stable native conformation of a protein determining a conformational change that leads to another stable conformation. (c) The ligand binds to a stable native conformation bringing to a more heterogeneous landscape of accessible conformations. Figure taken from Ref [10].

1.2 Tyrosine kinases as paradigms of highly dynamic structures

Tyrosine kinases (TKs) are very dynamic structures, perfect example of proteins with a highly multidimensional landscape, characterised by the adoption of several conformations, even if just a few are of relevance to this study. Indeed, the work presented in this thesis is focused on the study of the conformational dynamics associated with the "ac-

tive to inactive" conformational switch in the kinase domain (KD) of TKs. Hereafter, I will describe more in detail the KD structure, and in particular the structural features involved in the conformational changes and the allosteric network of the kinase.

1.2.1 Structure and conformational changes in tyrosine kinases

The kinase KD consists of about 250 residues and comprises a smaller N-terminal lobe and a larger, helix reach, C-terminal lobe (see Figure 1.3). At the interface between the lobes, lies the ATP (adenosine triphosphate) binding site formed by a number of highly conserved residues.

Tyrosine kinases have been observed in a variety of different conformations, active and inactive. Activation leads to ATP binding and is mainly controlled by conformational changes in three conserved structural motifs: the DFG-motif, the activation loop (A-loop) and the αC-helix[20](see Figure 1.3).

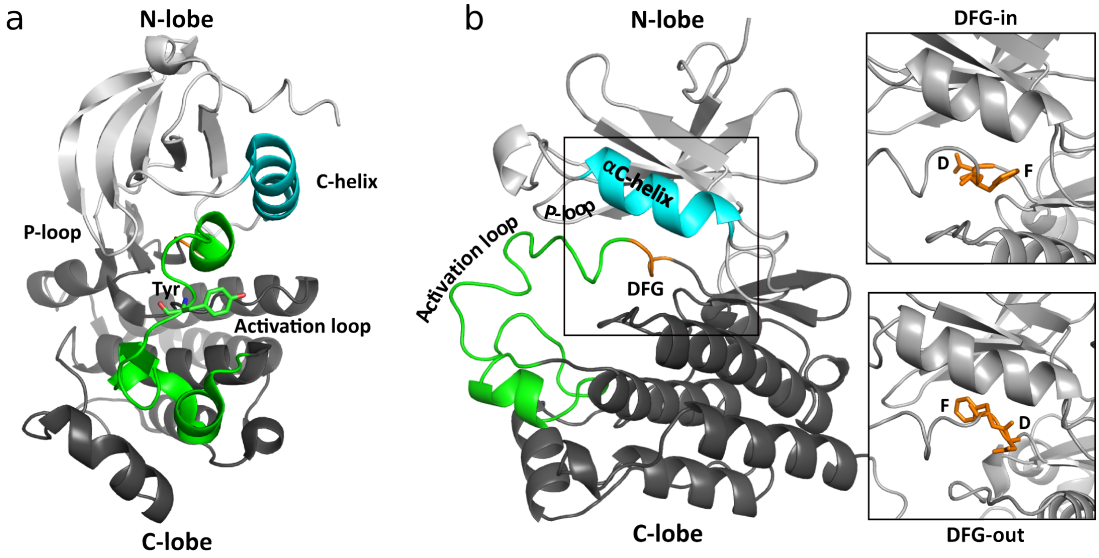

Figure 1.3: Front view of the Abl kinase domain (a) with highlighted in different colours the most important structural features and in sticks the tyrosine residue of the activation loop (in green), which phosphorylation activates the kinase. (b) Side view of Abl with a more detailed view of the DFG motif (in orange sticks) and of its two possible conformations, DFG-in and DFG-out.

The A-loop is a highly flexible region connecting the two lobes and can assume an ensemble of different conformations, going from a fully open to a closed form (see Figure 1.4), which hides the binding site (also called active site). When the A-loop is in an extended fully-open conformation, a specific tyrosine can be phosphorylated

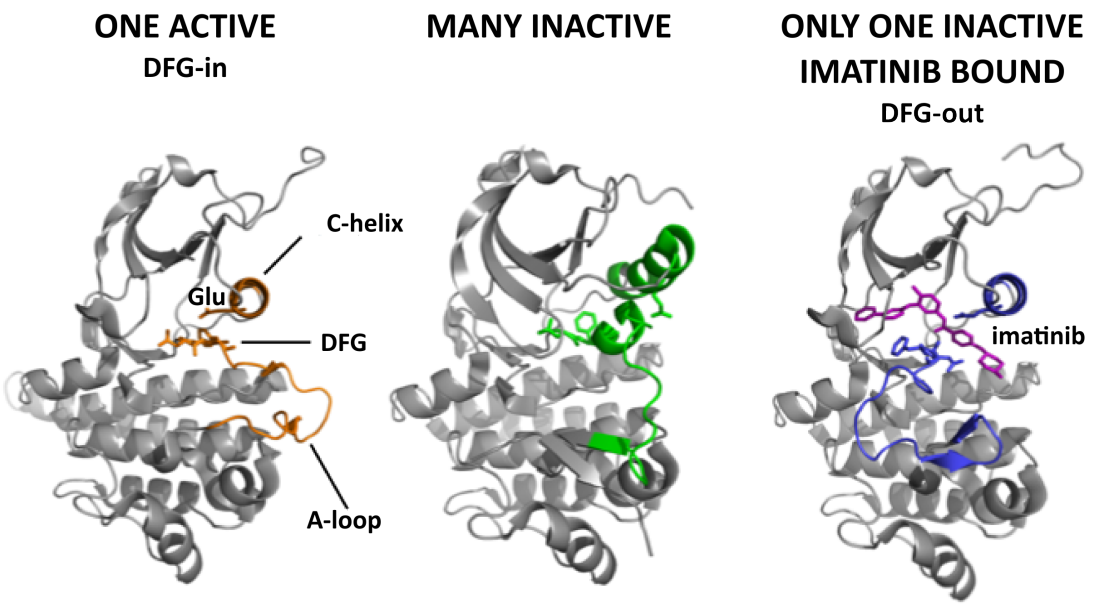

Figure 1.4: Schematic representation of the possible conformations a tyrosine kinase can assume. There is just one active conformation, with well defined structural characteristics to fulfil (DFG-in, A-loop open, αC-helix Glu-in) but many inactive intermediate conformations, with different arrangements the catalytic important elements can adopt. Among them just one is of interest in the context of this thesis work, and is the inactive DFG-out conformation in which the drug imatinib (in magenta sticks) binds to the kinase active site (A-loop closed, αC-helix Glu-in).

leading to kinase activation(see Figure 1.3, image a). On the contrary, all other possible A-loop conformations are considered inactive. The Asp-Phe-Gly (DFG) motif, at the N-terminus of the A-loop, is able to flip from an 'in' to an 'out' form (Figure 1.3 and Figure 1.4). In the active conformation, the Asp is pointing towards the ATP binding site (DFG-in or Asp-in) and the Phe away from it, occupying an hydrophobic pocket, while in the inactive conformation Asp and Phe exchange their positions (DFG-out or Asp-out). The DFG-in to DFG-out flip has been proposed to have a functional role in the release of the product of catalysis (adenosine diphosphate, ADP). Finally, the so-called αC-helix, localised in the N-lobe (Figure 1.3), has a glutamate that, both in the active and inactive conformations (Figure 1.4), is forming a salt-bridge with a lysine in the active site (αC-in or Glu-in). In the passage from active to inactive, the kinase adopts an intermediate conformation in which the glutamate of the αC-helix interrupts the salt-bridge with the lysine to form a new one, with an arginine of the A-loop, adopting a Glu-out position (see Figure 1.4).

1.2.2 Allostery in tyrosine kinases

The process by which a subtle perturbation in a region of a protein could be amplified and determine the motion of tens of residues, is still debated. TKs are the paradigm of signalling proteins and their activity is finely tuned by allostery. Even if the structural elements involved in the kinase allosteric network (see Figure 1.5) have been identified[17], their interplay is not completely understood. Moreover, the allosteric mechanisms within every kinase are unique and associated with protein conformational changes[21].

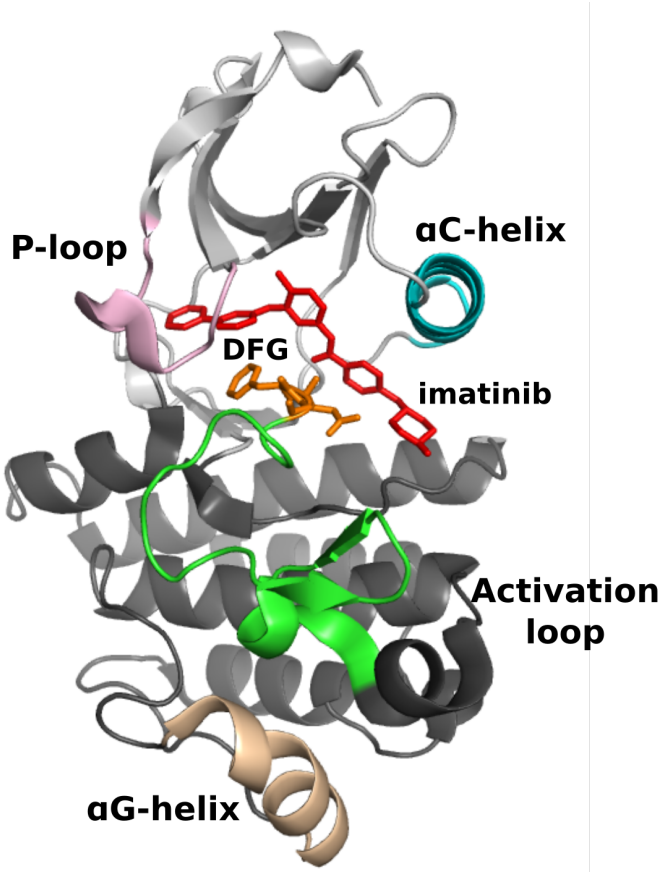

Figure 1.5: Regions of the kinase domain involved in the allosteric network of TKs. The different colours identify every element of the network: P-loop in pink, αC-helix in cyan, DFG motif in orange, A-loop in green and the αG-helix in wheat. Imatinib is shown in red sticks binding to the kinase active site.

Figure 1.5 reports the structural elements part of the allosteric network in most TKs, involved in the conformational changes of the kinase domain, mentioned in the previous section, and essential for ATP and drug binding. Some areas of the structure, shown in Figure 1.5, such as the αC-helix, the DFG motif, the A-loop are directly

involved in the conformational changes of the kinase domain, but others such as the P-loop and the αG-helix, experience a repositioning as a result of allosteric modulation.

Important for the structural integrity of the kinase core and also because considered part of the allosteric network[17] are: the catalytic spine (C-spine) and the regulatory spine (R-spine) (see Figure 1.6). These "spines" are two hydrophobic motifs, conserved among all kinases, and postulated to play a role in guaranteeing the kinase with an efficient and robust catalysis[22, 23, 24]. Described for the first time in the PKA kinase, both the C-spine and the R-spine span from the N-lobe to the C-lobe, creating a stable net of hydrophobic interactions, enabling the right positioning of the ATP and the substrate during the transfer of the phosphate group. The spines are fully assembled in the active kinase, while they are disrupted when the kinase is inactive. The C-spine (see Figure 1.6, in yellow) is completed by the adenine ring when ATP is bound to the binding site while, the R-spine is the first to be assembled as a consequence of A-loop phosphorylation (see Figure 1.6, in red). Moreover, the R-spine comprises catalytic important elements, such as the Phe of the DFG motif and a residue (Leu) of the αC-helix.

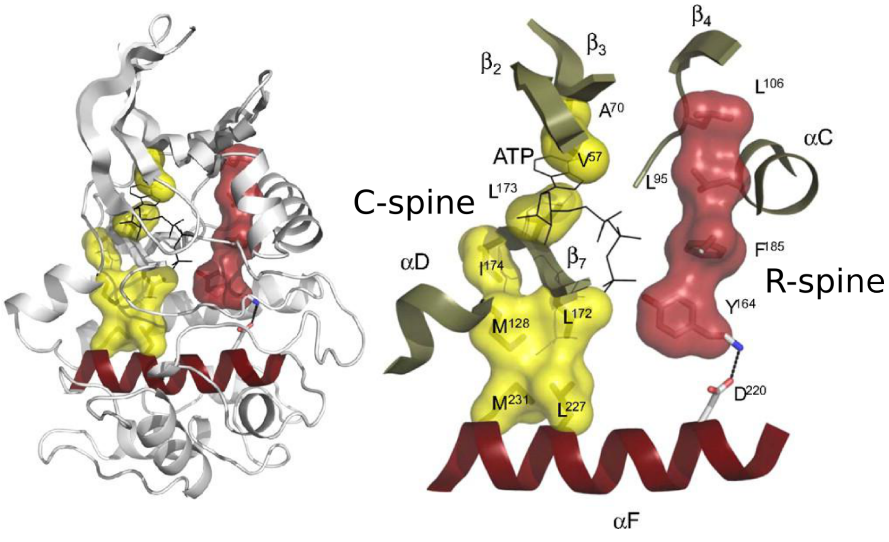

Figure 1.6: Protein Kinase A (PKA) structure where hydrophobic spines have been detected for the first time. View of kinase domain with the catalytic spine (C-spine) shown in yellow and the regulatory spine (R-spine) shown in red. Figure taken from Ref [23].

1.3 Tyrosine kinases and their deregulation in cancer

TKs were chosen as subjects of this study because of their great historical and biomedical importance as anti-cancer drugs targets[25]. In this section, I give first a general overview of the TK family, followed by a more detailed description of the TKs considered in the study.

1.3.1 The tyrosine kinase family

TKs are a large and diverse family of enzymes whose central role is to catalyse the transfer of a γ-phosphate from the ATP molecule to the amino acid acceptor on protein substrates[26] (see Figure 1.7). The phosphorylation event is crucial in the regulation

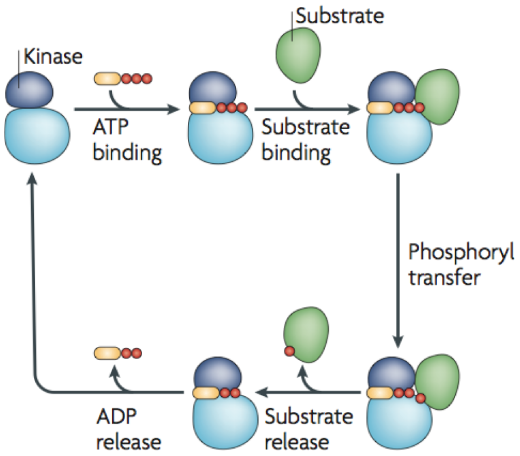

Figure 1.7: Catalytic cycle of a tyrosine kinase protein. Figure taken from Ref [26].

of cellular signalling and numerous biochemical reactions, under normal and pathological conditions[27]. Their deregulation is related to several human diseases, including diabetes, inflammation, cardiovascular diseases and cancer. The role of TKs in the regulation of cell cycle, cell proliferation and apoptosis makes them the most important targets for anti-cancer treatments[28, 29].

TKs are further divided into two families: the non-receptor tyrosine kinases, localised in the cytoplasm, and the receptor tyrosine kinases (RTKs), on the cell membrane. RTKs are constituted by three domains: an extracellular, a transmembrane and an intracellular catalytic domain. The binding of a ligand to the extracellular domain is often coupled to oligomerisation, causing conformational changes in the structure resulting in enzyme activation. The active enzyme is able to fulfil its biological function,

and thus to bind and phosphorylate cytoplasmic substrates[30]. The phosphorylation of intracellular proteins triggers a cascade of events for the transduction of the signal directly to the nucleus, where gene expression is regulated. In a similar way, non-receptor tyrosine kinases regulate several cellular functions by phosphorylating enzymes in response to events happening in the cytoplasm.

In pathological conditions, as in cancer, TKs become constitutively active, determining an uncontrolled phosphorylation of the kinase substrate, often leading to cell proliferation[31, 32].

Hereafter, it follows a brief description of the TKs considered in the study, to elucidate their cellular function and the result of their pathological deregulation. Src and Abl kinases were the main focus of this work because both targets of the successful anti-cancer drug imatinib (introduced later). Lck, Kit, Met and EGFR were later included in the work, and chosen for their high sequence homology (more than 45%) to both Src and Abl, and because imatinib targets.

1.3.2 The Src tyrosine kinase

Src is the homologue of the first discovered retroviral oncogene, v-Src. This gene, found in the Rous Sarcoma virus (RSV) is responsible for provoking uncontrolled mitosis of host cells, inducing sarcoma[33]. The enzyme is constituted by three domains:

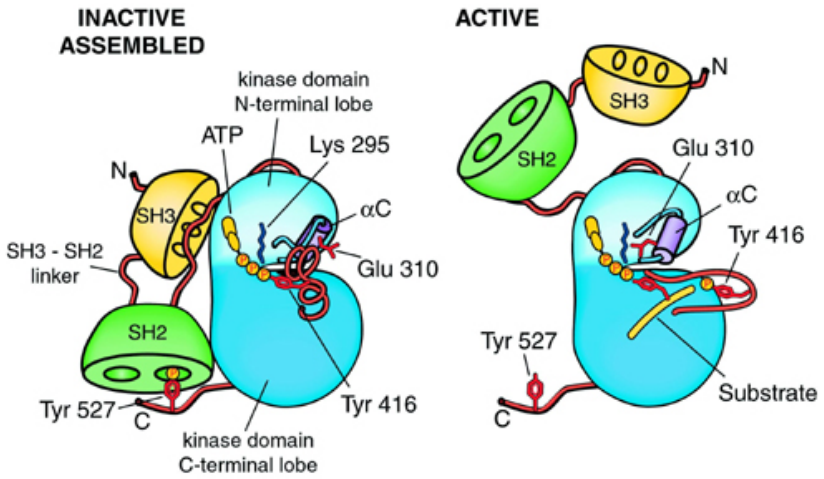

Figure 1.8: Structure of Src kinase. The full-length kinase is constituted by three domains: SH2, SH3 and kinase domain (KD, the only considered in this study). The different spatial orientation of the domains discriminates between the inactive and active forms, adoptable by Src. For the activation or inactivation of the full-length kinase, crucial are two sites of phosphorylation: Tyr527 of the C-terminal tail (phosphorylated in the inactive form) and Tyr416 of the activation loop (phosphorylated in the active form). Figure taken from Ref [34].

the SH2 and SH3 domains, localised at the N-terminus of the structure and connected, by a linker, to a larger catalytic domain (or kinase domain, KD), where the ATP binding site is present[35, 36, 37](see Figure 1.8).

The catalytically active form of Src, activated by the phosphorylation of Tyr416 on the activation loop (A-loop), is responsible for cell movement and proliferation. In the auto-inhibited conformation the phosphorylated Tyr527, localised on the C-terminal tail, interacts with the SH2 domain, while the linker with the SH3 leading to a closed conformation [34, 38, 39]. In cancer, Src is found in a constitutively active form, lacking the inhibitory phosphorylation site (Tyr527).

1.3.3 The Abl tyrosine kinase

Abl is a cytoplasmic tyrosine kinase encoded by the ABL1 gene placed on the chromosome 9. When a reciprocal translocation, between chromosome 9 and 22 occurs, the BCR-Abl fusion gene is formed. The union of the ABL1 gene (originally on ch9) with the BCR gene (on ch22) generates the so called "Philadelphia chromosome", most commonly associated with chronic myeloid leukemia (CML)[40].

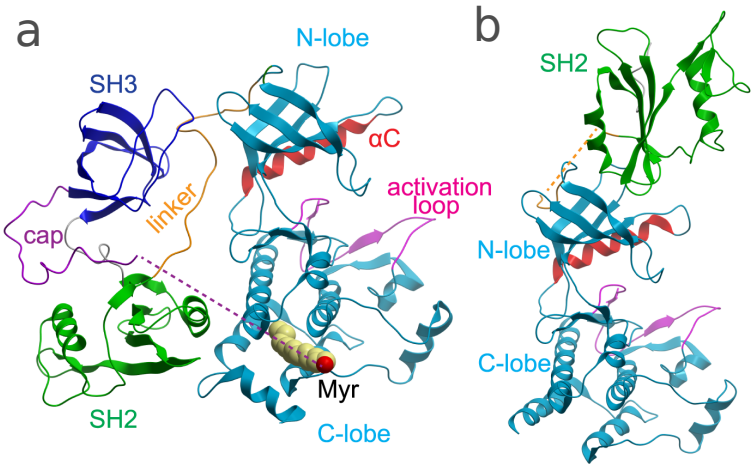

Figure 1.9: Cartoon representation of the Abl structure. (a) The full-length kinase is constituted by three domains: SH2, SH3 and kinase domain (KD, the only considered in the study). The kinase is in the inactive form with the myristate tail (Myr) docked to the hydrophobic pocket in the C-lobe. (b) Cartoon representation of the activated form of Abl, with the peculiar position adopted by the SH2 domain, on top of the kinase domain. Figure taken from Ref [41].

Abl is homologous to Src and the modular arrangement of SH2, SH3 and catalytic domain is common to both tyrosine kinases. The KD is connected at the N-terminus with the SH2 subunit by a linker, at which follows the SH3 (see Figure 1.9, image a). When activated by phosphorylation of the tyrosine on the activation loop (A-loop), the

kinase undergoes a large conformational change in which the SH2 domain moves at the top of the KD (see Figure 1.9, image b). Only inferences can be made about the position adopted by the SH3 in this process, since no crystal structures are available[42]. The protein encoded by the fusion gene (BCR-Abl) preserves Abl, with the exception of the myristate tail (Myr in Figure 1.9), a small N-terminal region before the SH3 subunit. The myristate tail docks to a hydrophobic pocket in the C-lobe of the protein, keeping it in an inactive form[43] (see Figure 1.9, image b). The loss of the myristate tail determines the deregulation of Abl kinase, resulting in the insurgence of CML.

1.3.4 Other tyrosine kinases of interest

Activating mutations of the genes codifying for TKs are associated with different types of cancer. For instance, Kit, in its mutant deregulated form, is responsible for the insurgence of gastrointestinal stromal tumours (GST)[44], EGFR of lung cancers[45], while Lck and Met are mostly involved in the metastatic process[46, 47].

Lck, the lymphocyte-specific protein, is found in specialised cells of the immune system, the T-cells. In normal conditions, it phosphorylates the tyrosine residue of the cluster of differentiation 3 (CD3) and of the T-cell receptor complex in the cell cytoplasm. After the phosphorylation event, a signalling cascade is started and it culminates with the intracellular release of Ca^{2+} ions, crucial for the activation of many signalling processes within the cell. Among them, the production of cytokines that promote lymphocytes proliferation[47].

The enzymes Kit, Met and EGFR are all tyrosine kinase receptors (RTKs). Kit is expressed on the surface of hematopoietic stem cells. When the ligand, the stem cell factor, binds to the protein it dimerises becoming active. The phosphorylation of second messengers in the cell determines the propagation of the signal, regulating proliferation and cell growth[44]. The Met kinase, also called hepatocyte growth factor receptor because of its ligand, the hepatocyte growth factor, is expressed on membranes of epithelial cells. When activated by the ligand, is involved in many important signal transduction pathways, such as the RAS pathway, the PI3K pathway and the STAT pathway among others. In normal conditions Met is responsible for mitogenesis and morphogenesis[46, 48]. Finally, the epidermal growth factor receptor (EGFR) is known to bind diverse growth factors. When a ligand binds, the formation of homodimers

or heterodimers is promoted and the signal transduction pathway initiated. EGFR is normally responsible for DNA synthesis and cell proliferation[49, 45, 50].

1.4 Tyrosine kinases as anti-cancer drugs targets

TKs have been the most attractive anti-cancer targets of the last decade[1, 51, 52]. Most TKs inhibitors are small ligands binding to the ATP binding site. One of biggest success of drug discovery[52], leading to a revolution in the treatment of chronic myeloid leukemia (CML) first, and gastrointestinal stromal tumours (GST) later, was the discovery of imatinib, commercialised with the name of Gleevec.

However, the different inhibitor activity (IC_{50}) of imatinib towards highly homologous TKs led to a major effort in the scientific community for the identification of the underlined mechanisms responsible for the observed differences.

Therefore in this section, after a brief overview of imatinib discovery and its use in therapy, I describe the several efforts and hypothesis formulated over the past 10 years, in the attempt to explain imatinib diverse activity towards TKs and its mechanism of binding.

1.4.1 Chromic myeloid leukemia and imatinib "promiscuity"

In 2001 the effective kinase inhibitor imatinib[53] was approved for the treatment of CML[54, 55]. Ninetyfive percent of patients, suffering for CML, are chromosome Philadelphia positive (Ph+) (see Section 1.3.3) and express the BCR-Abl fusion protein. CML is responsible for an uncontrolled myeloid cell proliferation in the bone marrow, with a subsequent abnormal accumulation of white cells in the blood[54, 55, 40].

Imatinib was developed to specifically target BCR-Abl[53, 56, 57]. The clinical success is due in particular to its stunning efficacy in the early stages of CML, where ninety percent of patients respond to the treatment, showing a complete remission [58, 59]. In stark contrast, 60% of patients, treated in an advanced stage of the disease, relapse in the long term due to the onset of drug-resistance [60].

Imatinib turned out to be not so selective[61] and to target, besides Abl, many other TKs (see Table 1.1). Imatinib inhibits effectively BCR-Abl[62], binding competitively to the ATP binding site of the kinase, but it is also successfully used in the treatment of gastrointestinal stromal tumors (GST), because of its serendipidous affinity for Kit[63].

TK	IC_{50} (μM)
Abl	0.037 [62]
Kit	0.3 [63]
Lck	9 [56]
Met	>100 [64, 65]
EGFR	>100 [66]
Src	>100 [66, 65]

Table 1.1: IC_{50} values of imatinib activity towards each of the targeted TKs. Lower IC_{50} values corresponds to a better imatinib activity (better kinase inhibition).

Although imatinib binds also to other kinases, such as Src, Met and EGFR among others, it does not effectively inhibit any of them[66, 60], despite the high sequence homology and x-ray structures showing the same binding mode to the active site[20] (see Figure 1.10). All the mentioned TKs have a 45% sequence homology with the KD of Abl, or even more if considering only the binding site, that is nearly identical. Puzzling is the case of Lck, quite effectively inhibited by imatinib[56][60], despite the high sequence identity with the KD of Src (70%).

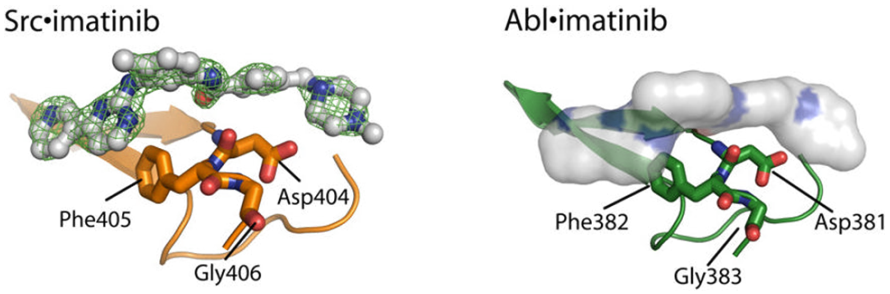

Figure 1.10: Binding mode of imatinib to Src and Abl obtained by x-ray crystallography. In sticks is shown the DFG motif adopting the "out" conformation. Imatinib adopts the same binding pose in both kinases. Figure taken from Ref [20].

Imatinib binds to the inactive (DFG-out) conformation of Src (see previous Section 1.2.1) with a 3000 times lower affinity than to Abl, resulting in a 4.6 kcal/mol penalty in binding energy[20]. This mismatch between drug affinity and structural similarity has been addressed numerous times in the past years but its molecular causes are still debated.

Initially, it was proposed that this difference was due to the inability of TKs, other than Abl, to adopt the DFG-out conformation[62, 67]. This hypothesis has soon been discarded, because of crystal structures showing imatinib-bound conformations of Src,

Kit and Lck[20, 68, 69]. In the light of these findings, it was proposed that the DFG-out conformation might play an important functional role and be accessible to many, if not most, kinases[70]. Subsequently, Seeliger *et al.*[20] ascribed the low imatinib activity towards Src to a difference in the conformation adopted by the P-loop region of Abl, which seems to rearrange upon drug binding (see Figure 1.11). This loop has been crystallised adopting a kinked conformation in the Abl-imatinib complex, while in Src and Kit it remains unchanged[20]. Since Kit is effectively inhibited by imatinib despite not showing a kinked P-loop, also this hypothesis was discarded.

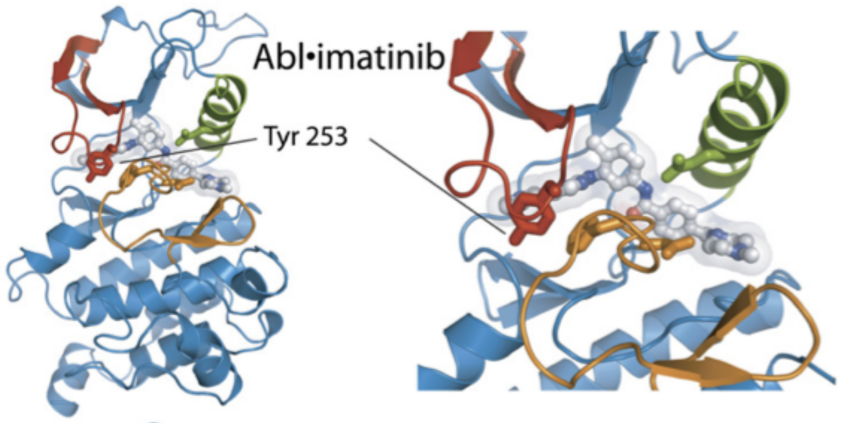

Figure 1.11: X-ray structure of Abl in the DFG-out conformation with imatinib bound. At a side, detailed image of the P-loop region. To be noted is the kinked conformation adopted by the P-loop (in red) and ascribed to reorientation of residue Tyr253 (in sticks). In Src and Kit instead of a tyrosine there is a phenylalanine. Coloured in green the αC-helix and in yellow the A-loop. Imatinib is shown in grey sticks. Figure taken from Ref [20].

At some point, also the protonation of residue Asp404, of the DFG motif, was hypothesised to influence the DFG-out stability, and therefore the binding of imatinib to Abl[70]. The DFG-out population in Abl was seen to increase around pH=6 and subsequently drop at pH$\simeq$7 (see Figure 1.12). Similarly, the K_{on} of imatinib binding to Abl drops around pH=7.5 (see Figure 1.12). Thus, at physiological conditions (cytoplasmic pH=7.2) the DFG-out population, able to bind imatinib, will be minimal.

Seeliger and co-workers also performed an extensive mutagenesis study[20], substituting residues of the binding site of Src into the corresponding ones of Abl, but they did not manage to show any substantial increase in the imatinib affinity for Src. Moreover, crystal structures and free energy calculations have shown that imatinib binds to the same residues, engaging in the same interactions in the binding sites of Src and Abl[71]. It follows that an answer should be looked for not in the structure of these pro-

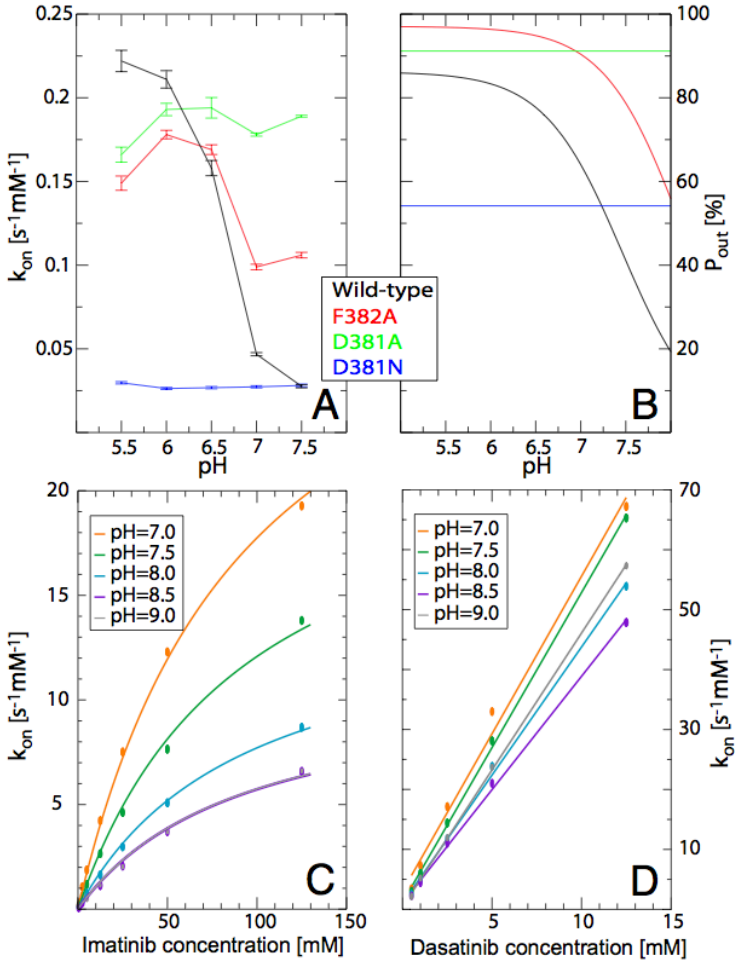

Figure 1.12: (A) pH dependence of imatinib binding to Abl wild-type (in black). (B) Variation of the percentage of the DFG-out population of Abl corresponding to different pH values (in black). (C) Binding rates of imatinib to Abl WT at different pH show high pH dependence for the binding. (D) Binding rates of dasatinib to Abl WT at different pH show weak pH dependence (dasatinib binds to the DFG-in conformation). Figure taken from Ref [70].

teins, but in the dynamics, kinetics and thermodynamics of the conformational changes they undergo.

1.4.2 Binding of Imatinib: conformational selection or induced fit?

The scaffold of imatinib is formed by a 2-phenylaminopyrimidine (see Figure 1.13 rings C and D), substituted with a benzamide-piperazine (rings A and B) and a pyridine (ring E). As shown in crystal structures of Src and Abl in complex with imatinib[20], the drug binds to the inactive (DFG-out) conformation of the kinase, adopting a bridge position above the DFG motif (see Figure 1.14), occupying both the ATP site, with

rings E and D, and the hydrophobic pocket (also called allosteric pocket) with rings A and B (see Figure 1.14). In Figure 1.13, the interactions imatinib makes, common to Abl, Src and the other TKs can be seen. Of particular interest are the ones it makes with the "gatekeepers" (Thr315 and Met319), Asp381 and Glu286, because they are all found to be mutated in patients showing drug resistance (see following Section 1.5.1).

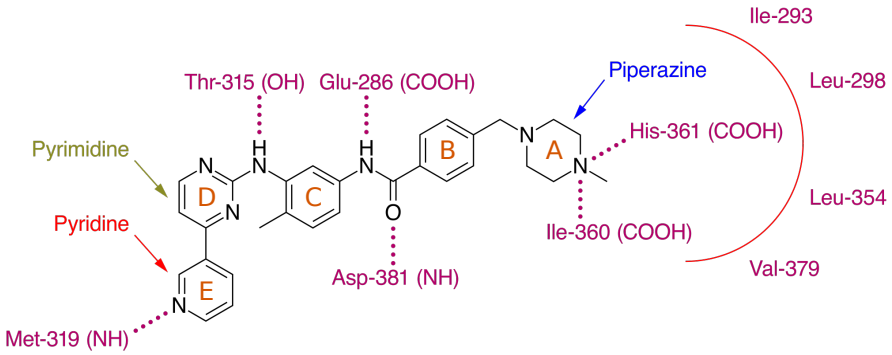

Figure 1.13: Structure of the drug imatinib. In dotted lines are shown the interactions it makes when bound to the kinase. The residues are common to Src and Abl. Figure taken from Ref [72].

The binding of imatinib to TKs seems to be influenced by different factors, primarily by the conformation of the DFG motif and the A-loop. The A-loop gains importance because when closed it is limiting the access to the binding site. Moreover, at the N-terminal region of the A-loop, there is the DFG motif which flip, from 'in' to 'out', it is known to be crucial for the binding to happen.

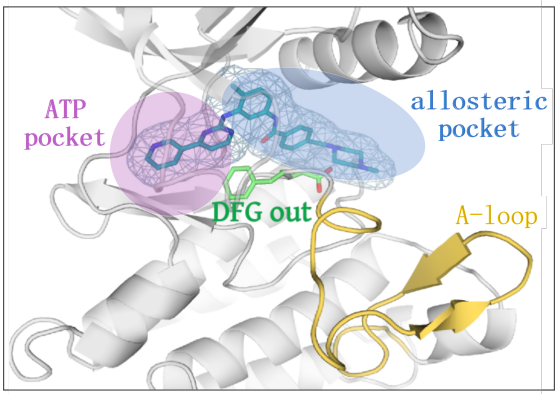

Figure 1.14: Imatinib binding to the active site of the kinase occupying both the ATP pocket and the allosteric pocket. When imatinib is bound, the kinase is adopting the DFG-out, inactive, conformation. Coloured in green is the DFG motif, in yellow the A-loop.

When the kinase is active (DFG-in), the A-loop is phosphorylated and it adopts an open conformation (see Section 1.2.1). It has been shown, in a recent work by Roux

et al.[73] on the Src KD, that when the A-loop is not phosphorylated, it can exists in a multi-state equilibrium of active and partially active conformations (see Figure 1.15). On the contrary, the phosphorylation locks the kinase in a catalytically active conformation (see Figure 1.15). Thus, phosphorylation stabilises the open conformation of the A-loop and the DFG-in as well, determining a shift in populations towards the active conformation, not targeted by imatinb[62].

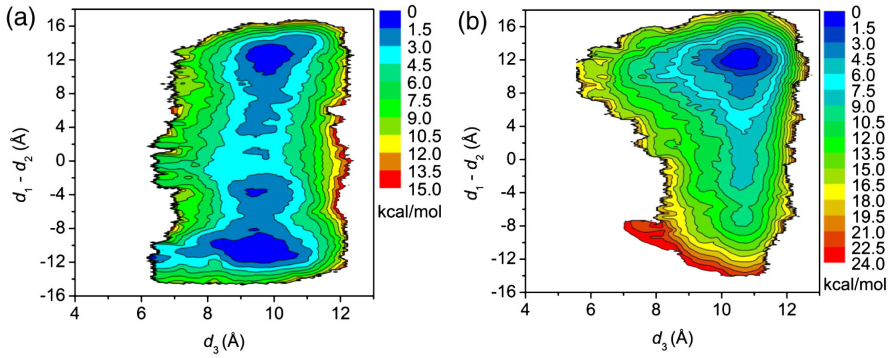

Figure 1.15: Free-energy surfaces of Src KD with A-loop not phosphorylated (a) and phosphorylated (b). The not phosphorylated surface (a) shows more energy minima while, the phosphorylated form (b) only one, corresponding to the open A-loop conformation. The method used in the study is the Umbrella Sampling enhanced method[74]. On the x-axis, the distance describing the A-loop closed (lower values) to open (higher values). On the y-axis, the subtraction of two distances describing the switch of the αC-helix from the Glu-in (higher values) to Glu-out (lower values) conformation. Figure taken from Ref [73].

Experiments performed on the phosphorylated and not phosphorylated forms of Abl, Src and Kit (see Figure 1.16), showed that the imatinib affinity for Src remains almost unaltered by the phosphorylation event[75] while, it is almost 30-fold reduced towards phosphorylated Abl[76, 20] and Kit[77]. This huge change in affinity towards both imatinib sensitive kinases, could be explained by considering that the druggable DFG-out conformation of Abl and Kit would be much more populated in the non-phosphorylated form of the two kinases, with respect of what would be for Src. Thus, the population shift towards the active conformation for Abl and Kit, due to phosphorylation, would result in a more drastic effect on the imatinib activity.

Also striking is the finding that phosphorylation seems to impair the binding only of imatinib and not of other anti-cancer drugs, such as dasatinib[76] and sunitinib[77]. The main difference between these drugs is the conformation of the kinase they target, for which we already know a major discriminant is the A-loop conformation. As discussed previously, imatinib binds to the inactive conformation (DFG-out, A-loop

closed) while, both dasatinib and sunitinib bind to the active conformation (DFG-in, A-loop open). Since the difference in activity of imatinib for TKs seems not to be so

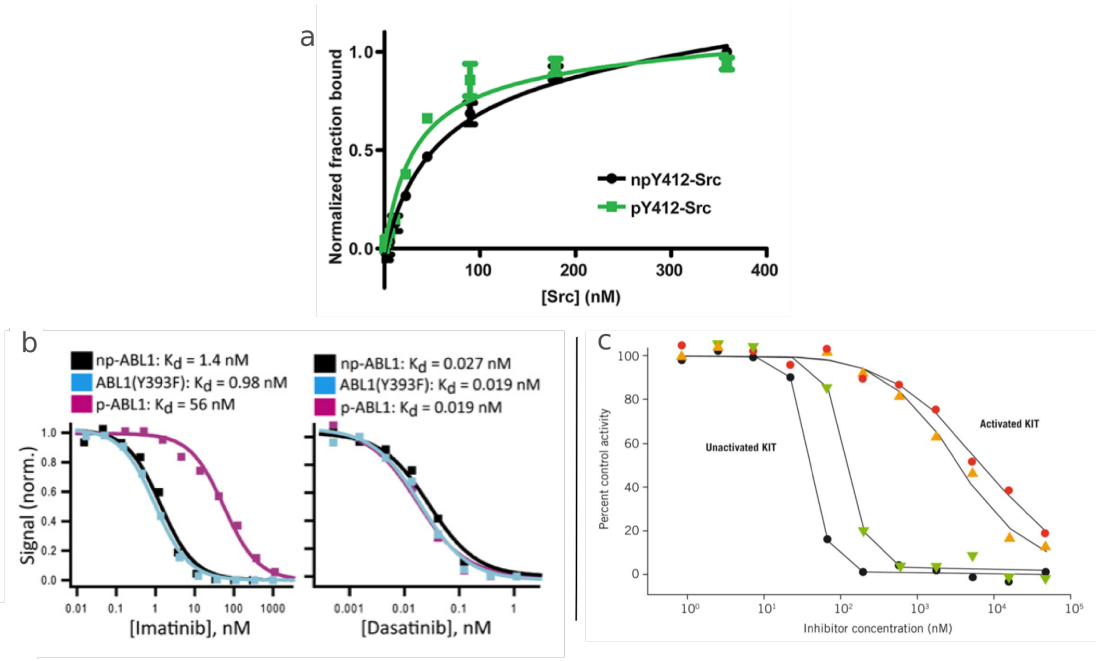

Figure 1.16: Response curves of drugs binding to the phosphorylated (p) and not phosphorylated (np) forms of Src (a) Abl (b) and Kit (c). Drug affinity has been tested with different methods in the three studies thus, values could not be directly compared. (a) Fraction bound of imatinib to Src "pY412" and "npY412". (b) Response curves of imatinib and dasatinib binding to Abl "p" and "np". It was also tested a mutant of Abl, in which phosphorylation has been impaired by mutating the tyrosine of the A-loop (Y393F). (c) Response curves of imatinib and sunitinib binding to Kit. The label "activated KIT" stands for phosphorylated while, "inactivate KIT" for not phosphorylated. Figure (a) taken from Ref [75], Figure (b) taken from Ref [76] and Figure (c) taken from Ref [77].

strongly influenced by individual structural features or sequence (see Section 1.4.1), more could be ascribed at the overall stability of TKs conformations, with a particular focus on the DFG motif and A-loop conformations.

The binding mechanism of imatinib to TKs has been long discussed, and both a conformational selection and an induced fit mechanism have been proposed[20, 78, 15, 79, 80].

In the works of Kern *et al.*[79, 80] it was proposed, on the basis of NMR and kinetics studies, that the higher affinity of imatinib for Abl might be due to an induced fit binding mechanism, related to the rearrangement of the P-loop region (kinked P-loop, discussed in Section 1.4.1 and shown in Figure 1.11). This rearrangement is supposed to happen after the docking of imatinib to the binding site. However, not to

forget is the crucial role played in first instance by the sampling, in both Src and Abl, of the druggable DFG-out conformation for the binding to take place, also underlined in other studies[78, 81].

In general, the hypothesis that both mechanisms could play a role in the binding of ligands to their targets has been postulated. It seems that the binding process could be described as a combination of two steps. A first step dominated by conformational selection mechanism, happening on larger time-scales, complemented by a second step of induced fit, more localised at the binding site[82, 83].

Moreover, a common statement in many of the works performed on Src and Abl, points to the fact that it is difficult to find a unique explanation for the different imatinib affinity, and that the free-energy penalty of binding could arise "from a distributed set of factors"[20].

1.5 Drug-resistance in tyrosine kinases

After the discovery of imatinib[53], a considerable effort has been devoted to the design and registration of new active, kinase targeting, anti-cancer drugs[84]. The demand for second-line drugs became even more urgent with the appearance of imatinib resistant cellular strains[85]. Indeed, the treatment with imatinib selects those cells harbouring mutations[86]. Still today, drug resistance remains the main drive for the development of new Abl inhibitors, because it is responsible for the relapses and the poor diagnosis of many chronic myeloid leukemia patients[65].

In this section, I describe the most common mechanisms of resistance and the advances in the molecular understanding of this phenomenon, together with the new drugs and strategies available to overcome imatinib resistance.

1.5.1 Imatinib and resistance

The numerous resistant mutations observed can be classified according to the degree of resistance they confer to the kinase inhibitor[87]. Mutations that can be overcome by increasing the dose of the drug are generally considered as conferring a lower degree of resistance, while those conferring a high degree of resistance are considered irreversible [65, 58].

Most imatinib resistant mutations are localised in the KD of Abl [88, 89] (see Fig-

ure 1.17) and some of them are directly involved in the binding. A classical mechanism, by which these missense mutations lower the affinity of the kinase for the inhibitor, is by abrogating specific molecular interactions, crucial for the stability of the binding[90] (see Figure 1.17). One of the most lethal mutations in Abl is the T315I, called the "gatekeeper' mutant[91] (see Figure 1.17), where the substitution of a threonine with an isoleucine eliminates an hydrogen bond crucial for the high-affinity binding of the drug.

Many other mutations (F317L, L387M, V379I, Y253F/H and F359V, see Figure 1.17), all lying in the active site, have a similar mechanism of action: they destabilise the binding or simply prevent it by steric hindrance[92]. Unfortunately, for several other resistance-causing KD mutations the mechanism of action is still unknown, because some of them are localised far from the active site or because are not directly involved in the destabilisation of the binding[88] (see Figure 1.17).

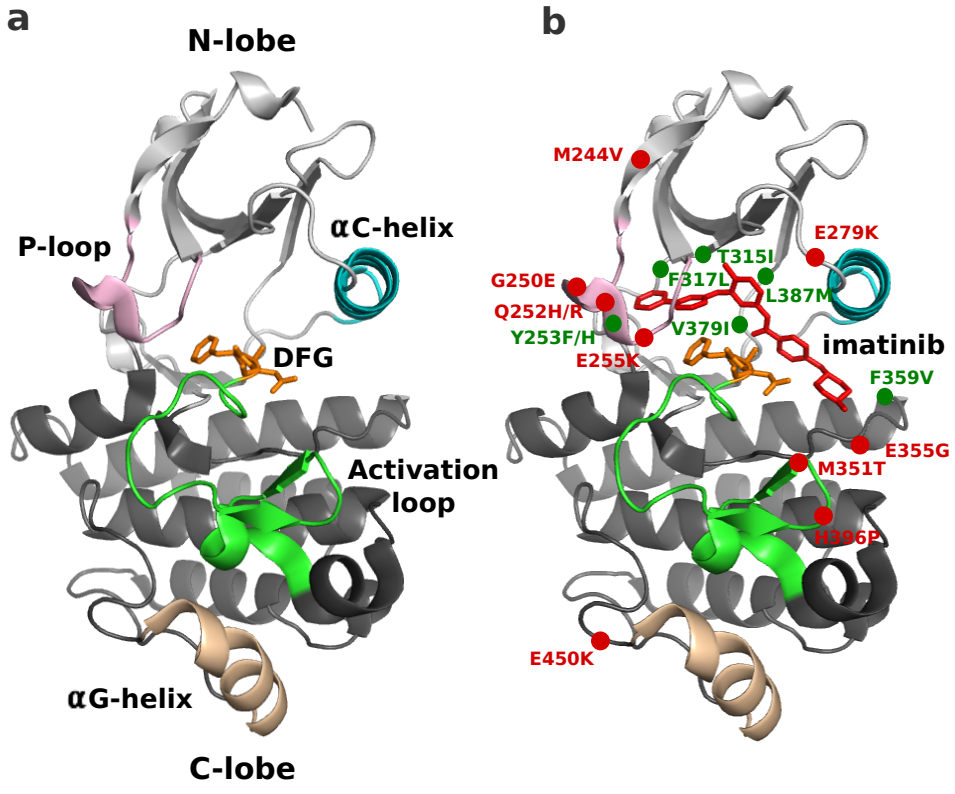

Figure 1.17: Abl KD (a) with resistant mutations highlighted (b). Green dots are those mutants with a known mechanism of action (*e.g.* elision of crucial bonds for the stability of the binding), while the red dots are the ones whose mechanism is still unknown[93].

Experimentally, it is known that the level of resistance to imatinib depends not only on the type of amino acid substitution[95], but also on the site where the mutation

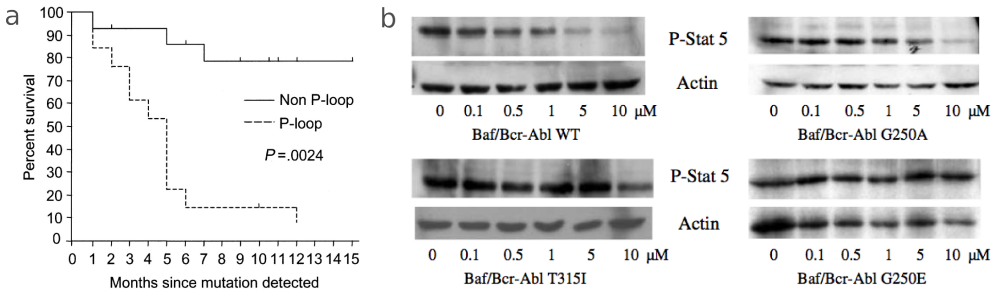

Figure 1.18: (a) Survival curves of patients expressing non P-loop and P-loop mutations. (b) Abl second messenger STAT5 phosphorylation levels for Abl WT and resistant mutants of Abl (T315I, G250E, G250A). T315I ("gatekeeper" mutant) and G250E (P-loop mutant) are not inhibited at high imatinb doses (above 5 μM, high degree of resistance). Figure (a) taken from Ref [94] and Figure (b) taken from Ref [95].

arises. As an example, patients expressing imatinib resistant mutations in the P-loop region (G250E/A, Q252H, Y253F, E255V/K), have on average worse survival rates[94] (see Figure 1.18). In particular, the study of Mahon *et al.*[95] shows very well how the substitution glycine to glutammate (G250E) is much more potent (higher phosphorylation rates of the substrate STAT5[96] of BCR-Abl) than that of the glycine to alanine (G250A, see Figure 1.18, image b). It seems clear though that even a single amino acid change could be critical for drug activity[97].

In the light of these findings and of the previously described binding mode of imatinib (adopting a bridge position over the DFG motif when in DFG-out, see Section 1.4.2), the interplay among sequence, site of the mutation and conformational dynamics should not be underestimated.

Imatinib resistance has clear mutational hotspots within the TK structure[98]. Mapping them could be useful in ranking the importance played by every region of the structure, and its relation with the binding and the affinity of the drug. Bioinformatics studies have revealed that many oncogenic mutations occur in conserved positions along the KD[98] and that these hotspots are recurrent and shared by several TKs[99]. The conservation of structure and often of sequence of mutations, suggests a conserved mechanism of drug resistance among kinases[1].

Recent computational studies, on oncogenic Abl, EGFR, B-Raf and FGFR[100, 101, 102, 103] kinases, have shown that, most of the times, mutations have an impact that goes beyond the mutated site, leading to wider allosteric changes in TKs (see Figure 1.19). They could determine a mutation-induced modulation responsible for confor-

mational selection and population shifts[100, 98, 101]. For example, mutations could be responsible for "cracking points" in proteins, enhancing partial unfolding events. The "cracking" theory ascribes to local unfolding to act as a catalyser of transitions in proteins, making possible global conformational changes[104, 105, 106].

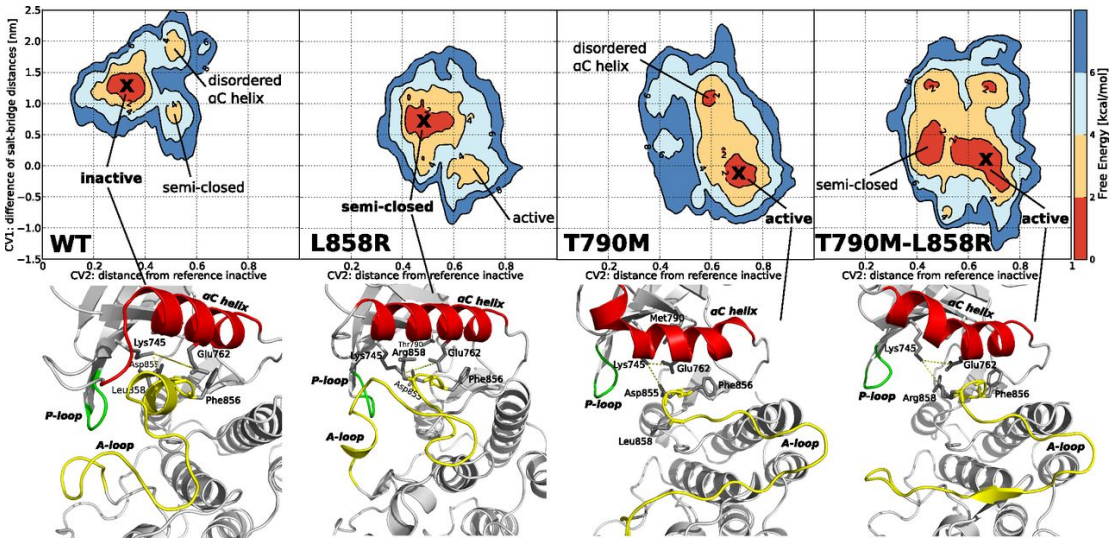

Figure 1.19: Free-energy surfaces of EGFR WT and of three EGFR oncogenic mutants (L858R, T790M and T790M-L858R). The mutations are responsible for a shift in the conformational space sampled by the kinase, leading to the stabilisation of the active conformation (open active-like A-loop and αC-helix Glu-in). Figure taken from Ref [100]

All major kinase inhibitors used in therapy educe drug resistance responses[107, 108, 109, 90]. A plausible mechanism for the mode of action of these resistant mutants, would be the destabilisation of the protein conformation with a high affinity for the drug (*e.g.* inactive, DFG-out conformation, in the case of imatinib) or the stabilisation of low affinity ones[89, 110, 111] (*e.g.* kinase active conformation, see Section 1.4.2). Detecting the features responsible for the selectivity and the resistance to imatinib could provide crucial information for the rational design of more effective and less toxic anticancer drugs.

1.5.2 Alternative strategies and drugs to overcome resistance

While in their catalytically competent (active) conformation TKs share a common fold, inactive conformations are numerous and highly diverse. The high structural homology among kinases complicates the search for selective inhibitors. Targeting inactive conformations is one of the strategies used to achieve a restricted selectivity[28].

Four classes of kinase inhibitors are successfully used in therapy, the main differ-

ence being their binding mode[112, 51]. Imatinib is known as a type II drug. Type II drugs bind to the inactive (DFG-out) conformation of the kinase, occupying both the ATP site and the adjacent allosteric pocket, that is freed due to the rearrangement of the A-loop and the peculiar Asp-out position adopted by the DFG motif[113] (see Figure 1.20). Due to the insurgence of imatinib resistance, alternative inhibition strategies have been developed along the years[1, 109, 114, 90]. Variability in the scaffolds or in the mode of binding of drugs have been used to overcome resistance. This is the case of nilotinib and sorafenib, type II drugs as imatinib but with slightly modified structures. On the contrary, dasatinib, sunitinib and lapatinib are type I drugs, that bind to the active (DFG-in) conformation of the kinase, occupying only the ATP site, right below the P-loop region[115, 108] (see Figure 1.20).

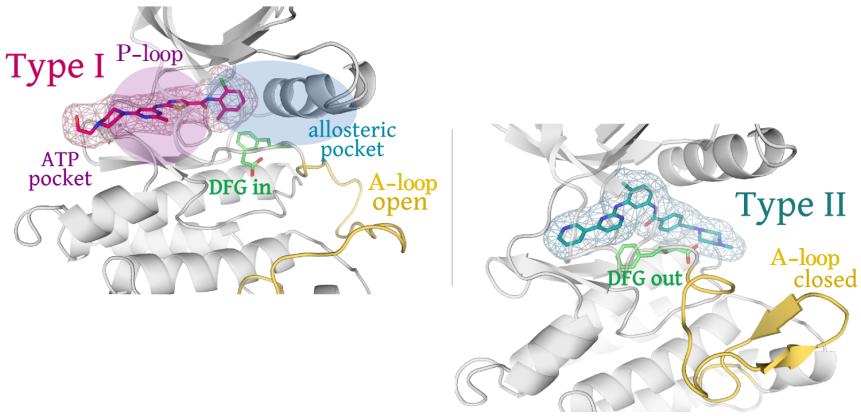

Figure 1.20: Binding modes of the two major classes of kinase inhibitors, Type I and Type II. Type I drugs bind to the ATP pocket, below the P-loop region, type II drugs occupy both the ATP and the allosteric pockets, adopting a bridge position over the DFG motif at the N-terminal of the A-loop.

More recently, allosteric kinase inhibitors have been introduced in therapy, used alone or in association with active site binders[85]. These drugs bind to remote sites of the structure, causing conformational changes responsible for the kinase inactivation. In the case of Abl, particularly exploited is the myristate pocket, located in the C-lobe (see Section 1.3.3). Drugs such as GNF-2 and GNF-5 are the most renown myristate binders[85, 15]. Their use in association with imatinib, has been successful in many cases of resistance, including the "gatekeeper" mutant T315I[85, 109].

Finally, there are the covalent inhibitors, that bind irreversibly to the ATP site. They could be either peptides, analogues of the biological substrate of the kinase, or small molecules specifically designed to target a cysteine residue present in the active site of most kinases[116, 117].

In the search for efficient TKs inhibitors, pivotal would be the understanding of the complex interplay among sub-μs dynamics, conformational changes and allostery in these targets.

Chapter 2

Materials and Methods

In this thesis I have used a combination of enhanced sampling atomistic molecular dynamics (MD) simulations and biophysical experiments to characterise the conformational dynamics and the flexible binding of ligands to a set of oncogenic tyrosine kinases.

Notwithstanding a large increase in the available computer power and much faster MD codes, most ligand binding and conformational changes events are still out of reach of atomistic simulations. This is due to the limited time-scales that can be simulated with atomistic MD that, even on the largest supercomputers[118], range from hundreds of ns to a few μs.

Since the aim of this thesis was the study of events (such as large-scale conformational changes in tyrosine kinases, TKs) that happen on longer time scales (ms), the use of enhanced sampling algorithms was required. The algorithm of choice was parallel tempering metadynamics (PTmetaD)[119], found to be very effective in sampling large scale conformational changes in TKs, and allows an accurate free energy reconstruction. I also used advanced analysis techniques to analyse the allosteric network connecting remote sites to the binding pocket. The expression and purification of the proteins to perform the biophysical experiments also required a lengthy trial and error procedure to optimise the yield. These are described below together with ITC approaches.

2.1 Computational Methods

2.1.1 Molecular Dynamics

Molecular dynamics (MD) simulations were used to compute the equilibrium properties of classical many-body systems, whereby a collection of N mutually interacting particles obey the laws of classical mechanics[120]. The time evolution of a system of N atoms, with Cartesian coordinates $\boldsymbol{r}_i$, can be simulated by integrating Newton's equation of motion over time:

$$\boldsymbol{F}_i(\boldsymbol{r}_1,...,\boldsymbol{r}_N) = m_i\ddot{\boldsymbol{r}}_i \tag{2.1}$$

where m_i is the mass of atom i, $\ddot{\boldsymbol{r}}_i$ the acceleration of atom i and $\boldsymbol{F}_i = -\nabla_{\boldsymbol{r}_i}V$ is the force acting on each atom, and V the potential (from the molecular mechanic force field). In practice, the previous set of coupled equations are solved numerically using the velocity-Verlet integrator[121, 122, 123]. Every time step, forces are calculated and new atoms positions updated in a trajectory file[120]. To assure the conservation of energy and the long term stability of the simulation, the time step needs to be ($\simeq$ 10 times) faster than the fastest vibrations (bond stretching), typically limiting the time step in the range of 1-2 fs.

The force field describes the potential energy of the system under analysis[124, 125, 126]. The most used force fields in classical MD simulations are still mainly based on the formulation by Lifson[127] originally devised in the 60s. This functional form is the result of two main contributions, the bonded and non-bonded interactions, defined as follows:

$$U = \underbrace{U_{bonds} + U_{angles} + U_{dihedrals}}_{\text{bonded interactions}} + \underbrace{U_{VdW} + U_{electrostatic}}_{\text{non-bonded interactions}}$$

$$= \sum_{\text{bonds}} \frac{1}{2} K_r (r - r_{eq})^2 + \sum_{\text{angles}} \frac{1}{2} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} K_n [1 + \cos(n\phi - \delta)] \tag{2.2}$$

$$+ \sum_{\text{non bonded pairs}} 4\varepsilon_{ij} \left[\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - \left(\frac{r_{ij}^0}{r_{ij}} \right)^{6} \right] + \sum_{\text{partial charges}} \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}}$$

where K_r, K_θ and K_n are force constants, r_{eq} and θ_{eq} are the equilibrium bond length and bond angle respectively. ϕ is the dihedral angle with δ being the phase angle. ε_{ij} is the Lennard-Jones well depth and r_{ij}^0 is the distance at which the value of the Lennard-Jones potential is zero, while the minimum is located at $2^{1/6} r_{ij}^0$. q_i and q_j are the partial atomic charges, r_{ij} the distance between them and ε_0 is the dielectric constant.

This simple functional form of the force-field neglects important contributions, such as the polarizability, but still it provides a good compromise between speed and accuracy. The most recent non-polarizable protein force-fields such as Amber ff99SB*ILDN[128, 129] and CHARMM22*[130, 131] are remarkably accurate, thanks to a significant re-parametrization effort, based on the comparison of the simulations to high-level ab-initio calculations and NMR experiments[132]. These force-fields are accessible through widely used molecular dynamics packages such as: AMBER[133], GROMACS[134], ACEMD[135], CHARMM[136], NAMD[137] and Desmond[138]. Most of these codes are parallel and can take advantage of graphics processing units (GPUs).

Still, computing the non-bonded interactions (electrostatic and van der Waals), for systems containing 50,000 or more particles, is time consuming. Even using a cut-off or more advanced approaches to reduce the computational cost of computing the electrostatic potential, such as the Particle Mesh Ewald (PME) method[139].

Integrating the equations of motion more than a billion times (equivalent to 2 μs, if the time step is set to 2 fs) to sample conformational changes, is prohibitively expensive even on a large-scale super-computer. Thus the need for enhanced-sampling algorithms to address this "timescale" problem is evident.

MD simulations are usually performed within the canonical (NVT) or isothermal-isobaric (NPT) ensembles, that maintain a constant number of particles (N), temperature (T), volume (V) or pressure (P), respectively[120]. In the NVT ensemble the system exchanges energy with a thermostat. Several are available, the more popular are: the Berendsen thermostat[140], the Nosé-Hoover thermostat[141], the velocity-rescale thermostat[142] and the Langevin thermostat[120]. While, the most commonly used barostats are: the Berendsen barostat[140] and the Parrinello-Rahman barostat[143].

Due to its robustness, fast equilibration times and simple implementation the velocity-rescale thermostat was used in the work presented in this thesis, while the Parrinello-Rahman algorithm was used to equilibrate the system pressure. Moreover, also the Langevin thermostat and the Berendsen barostat were used in simulations performed with the ACEMD software.

2.1.2 Enhanced sampling algorithms and free energy calculations

The "timescale" problem is currently one of the main limitations of MD simulations. For this reason many algorithms to enhance the sampling and reconstruct free energies have been proposed. Broadly speaking, these algorithms can be classified in four categories[144, 145, 146]:

- Methods to enhance the sampling as a function of one or a few predefined collective variables (CVs) (*e.g.* Umbrella Sampling[147, 74], Thermodynamic Integration[148], Metadynamics[149, 150, 151] and its derivatives such as Parallel Tempering Metadynamics[119], Multiple Walkers Metadynamics[152], Bias Exchange Metadynamics[153], Path Collective Variables[154], Steered MD[155], Reconnaissance Metadynamics[156].).

- Methods in which the phase space is explored simultaneously at different values of the temperature (*e.g.* Parallel Tempering[157], Parallel Tempering Metadynamics[119]).

- Methods aimed at exploring the transition mechanism and constructing reactive trajectories (*e.g.* Path Collective Variables[154], Steered MD[155], String method[158, 159], Thermodynamic Integration[148]).

- Methods for exploring the potential energy surface and localising the saddle points that correspond to the transition states (*e.g.* Metadynamics[149, 150, 151], Multiple Walkers Metadynamics[152], Transition Path Sampling[160], Markov State Models[161]).

Extensively applied in this thesis, for the reconstruction of the DFG flip conformational change (DFG-in to DFG-out), was the Parallel Tempering Metadynamics method[119] combining a CV-based method to Parallel Tempering.

2.1.3 Metadynamics and Parallel-Tempering Metadynamics

One of the most effective and widely-used family of enhanced-sampling algorithms is Metadynamics (metaD)[149, 150]. Metadynamics allows for a fast and accurate exploration of the free energy surface (FES). The FES is the reconstruction of the system free-energy as a function of collective variables (CVs). Thus, metaD is somewhat similar to CV-based approaches such as Umbrella Sampling[147, 74]. It requires the identification of a set of CVs that are able to describe the process of interest (*e.g.* conformational change, binding of ligand or a chemical reaction). It can treat several CVs (usually up to 4) simultaneously, and can be used both for reconstructing the free energy and for accelerating rare events.

The dynamics in the space of the chosen CVs is enhanced by a history-dependent potential constructed as a sum of Gaussians (or rational functions) centred along the trajectory followed by the CVs. The CVs need to describe the reaction coordinate and include all the slow degrees of freedom relevant to the reaction. When all the free energy basins for the system have been explored, the free energy surface can be reconstructed by integrating the bias.

It has been recently proved that the well-tempered implementation of Metadynamics provides an unbiased estimate of the free energy of the system projected onto predefined CVs[162]. However, the convergence time depends critically on how optimal are the CVs. Diverse and highly versatile CVs have been implemented over the years in the open source PLUMED plug-in package, increasing the success of the method[163].

The $V_{bias}(s)$ is added at defined time-intervals (τ_G) as Gaussians, with a given height (w) and width (σ), chosen by the user. The main aim of metaD is the filling of energy basins with 'tailored' Gaussians (w and σ), enabling the system to overcome high energy barriers and to explore a wider conformational space (see Figure 2.1).

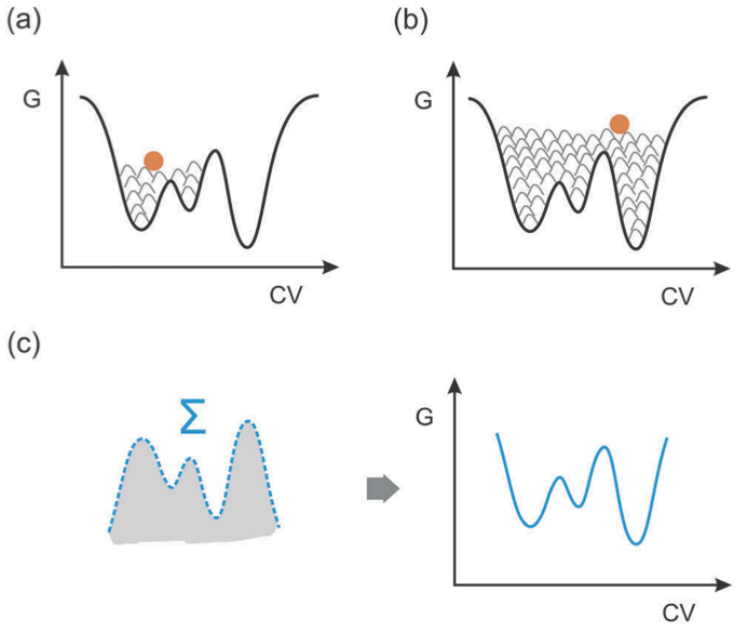

Figure 2.1: Gaussians are added along the CV space to help the system overcoming energy barriers during the metaD simulation (a). When the entire conformational space is filled with Gaussians (b), the added bias is summed (c) and the FE profile reconstructed (d). Figure taken from Ref [164].

In the well-tempered metaD[151], the height w of the Gaussians is rescaled over time. w tends to zero with the increase of the simulation time, as $\propto \frac{1}{t}$, leading to a fast convergence and a correct free energy reconstruction:

$$w = \omega e^{-\frac{V(s,t)}{\Delta T}} \tau_G \tag{2.3}$$

where ω is the initial deposition rate of the Gaussians and τ_G is the time interval for the deposition. Over time, the variation of the bias potential will be so small, that the system will be considered at equilibrium, and converged to the right free energy value:

$$V(s,t \to \infty) = -\frac{\Delta T}{\Delta T + T} F(s) \tag{2.4}$$

where $\frac{\Delta T}{\Delta T+T}$ is the bias factor. ΔT could be tailored by the user in order to accelerate the crossing of energy barriers and the exploration along the s space.

Well-tempered metaD solved many of the convergence problems of standard metaD. Still the main limitation remains the strong dependence on the choice of the CVs. Indeed, forgetting a relevant "slow" variable could lead to extremely slowly converging free energy reconstructions, making the approach impractical.

The Parallel Tempering Metadynamics method (PTmetaD)[119], used in this study, was developed to minimise the convergence problems due to the choice of the CVs. PTmetaD is a combination of well-tempered[151] Metadynamics (metaD)[149, 150] and Parallel Tempering (PT)[157] methods. The combination of the two (PT+ well-tempered metaD) permits to overcome the weaknesses of both methods, and to enhance effectively the sampling of conformational landscapes, to compute fully converged free energy.

PT helps accelerating MD simulations, running replicas of the same system at different temperatures[157], resembling the Replica Exchange Molecular Dynamics (REMD) method[165]. The replicas of the system running at higher temperatures manage to cross FE barriers more easily than it would do at lower temperatures thus, exploring a wider conformational space. Every MD simulation runs in parallel at a specified temperature (T_1<T_2<...<T_N), for a total of N replicas. Periodically, an exchange between the coordinates of two neighbour replicas (T_1 and T_2) is attempted (see Figure 2.2), and accepted with a probability:

$$P(\mathrm{acc}) = \min 1 \left\{ \exp[(\frac{1}{\mathrm{k_B T_2}} - \frac{1}{\mathrm{k_B T_1}})(\mathrm{U_2 - U_1})] \right\} \tag{2.5}$$

where k_B is the Boltzmann constant for the two replicas of the system, T is the temperature and U is the potential energy. The acceptance for the exchange follows the Metropolis criterion and depends on the potential energy of the two conformations (U_1 = U_2). The coupling of systems at different temperatures assures the enhanced sampling along all degrees of freedom. However, to improve the sampling of specific conformational changes of interest, PT is associated with well-tempered metaD.

The PTmetaD free energy surface (FES) is finally obtained by reconstructing the FE of the system at a specific temperature T. However, in the summation of the added Gaussians, a ΔT (bias factor) increase has to be considered, to account for the altered

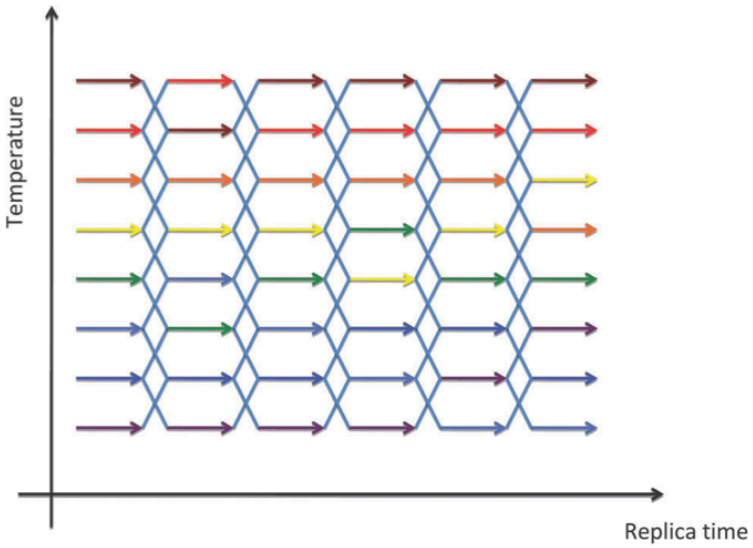

Figure 2.2: Schematic example of the exchanges happening among replicas at different temperatures during a Parallel Tempering simulation. Figure taken from Ref [166].

probability distribution in the s space. The choice by the user of w (height of the Gaussians), σ (width of the Gaussians) and τ_G (time interval for the deposition of Gaussians) is decisive for the accuracy of the sampling, and of the FE reconstruction.

As mentioned before, critical for the set up of the metaD calculation, is the choice of CVs that best describe the process to investigate. Even if with metaD is possible to treat several CVs simultaneously, limiting the number of variables is preferable, in order to prevent poor computational performances. In the light of this, the use of PT in combination with metaD, provides for the neglected CVs, thank to the enhanced sampling assured at higher temperatures.

2.1.4 Classical MD simulations details

In this thesis, classical MD simulations of several tyrosine kinases (TKs) were performed, considering the kinase catalytic domain (KD). The KD counts 260 amino acids. Crystal structures of the KD, adopting the DFG-in conformation, were retrieved from the Protein Data Bank, PDB entries Src:2SRC, Abl:2G1T, Kit:1PKG, Lck:3KMM, Met:2WGJ and EGFR:2ITW.

Missing residues were added with the Modeller software[167], according to the respective Uniprot sequences. The Amber99SB*-ILDN[128, 168, 169] force field was employed and the topology files were created with AMBER[133], using TIP3P[170] water molecules in a cubic box, with periodic boundary conditions and a minimum dis-

tance of 10 nm between the protein and the boundaries. The system charges were neutralised, adding the proper number of positive (Na+) or negative (Cl-) ions. The systems of all TKs were minimised with 10000 steps of conjugated gradient and equilibrated using ACEMD[135] in the NPT ensemble for 10 ns, using a Berendsen barostat[140] at 1 atm. The temperature was kept at 310K by a Langevin thermostat[120]. Classical MD simulation were carried out for all the systems in the NVT ensemble with a time step of 4 fs. The runs of Abl, Src and Kit were extended to a total length of 1 μs, for testing purposes, while runs of Lck, Met and EGFR were 400 ns long. All MD simulations were performed with the ACEMD program[135] running on a server with four Nvidia Fermi GPUs. The van der Waals interactions were cut off at 1.0 nm, and the long range electrostatic interactions were calculated by the PME algorithm[139] with a mesh spaced 0.1 nm.

The same MD setup was employed in the simulations of the resistant mutants of Abl (G250E, E279K, H396P, E450K and T315I) and of the engineered mutants of Src (Q275A/P299Q/Q420E, Q275G/P299E/Q420A, Q275A/P299Q/V461S, Q275A/P299Q/Q309A, E415G/Y416G/V461S, E415G/Y416G, M302G and αC chimera). The mutations were modelled in-silico using the Modeller software[167]. All MD simulations of the mutants last for 400 ns.

2.1.5 Trajectory analysis

Root mean square deviation analysis

The root mean square deviation analysis (RMSD) was used to monitor the behaviour of the structure during the MD simulation. The displacement of the system with respect to the initial, equilibrated, structure gives an insight of the stability and changes undergone by the system. The analysis was performed using the g_rms tool of GROMACS[134] and considering the atoms of the backbone.

Root mean square fluctuation analysis

To study sub-μs dynamics in TKs, root mean square fluctuation analysis (RMSF) was performed using the tool g_rmsf of the GROMACS[134] package. The fluctuations of every Cα atom, with respect to a reference structure, were averaged along the trajectory. The first 50 ns of the MD trajectory were discarded, based on the RMSD analysis, because considered equilibration time. To remove the effect of large scale conforma-

tional changes, the analysis was calculated on overlapping windows of 100 ns each and averaged, taking as reference the initial structure after equilibration time.

Principal component analysis

To investigate the essential dynamics and the dominant motions in TKs, principal component analysis (PCA)[171, 172] was performed using GROMACS[134]. The covariance matrix of the MD trajectory was calculated with g_covar, considering the protein backbone. The eigenvectors and eigenvalues, generated by the diagonalisation of the covariance matrix were analysed with the tool g_anaeig. To consistently compare the conformational space visited by every kinase with respect to Abl (chosen as reference of a flexible kinase in the study), the trajectories of all TKs and resistant mutants of Abl were projected on Abl first and second eigenvectors.

Solvent accessible surface area analysis

The solvent accessible surface area analysis (SAS)[173, 174] was used to compare the dimension of the kinase binding site in the DFG-in and DFG-out conformation of Src and Abl.

Short classical MD simulations (20 ns) were performed using as starting structures the most representative DFG-in and DFG-out conformations, extracted from the free energy minima of the PTmetaD calculations and then clustered using g_cluster tool.

The SAS analysis was performed using the GROMACS[134] tool g_sas, based on the double cubic lattice method discussed in Ref [174]. The binding site cavity of both kinases was defined by residues: 275-279, 296-340 and 391-412 (PDB 2SRC numbering). Since the data point of the MD trajectory would not be independent, the SAS analysis was calculated on 100 structures randomly chosen for both DFG-in and DFG-out simulations. The process was repeated 100 times and the SAS averaged afterwards. Default values were used for the number of dots per sphere (24) and the water probe radius (0.14 nm).

Subsequently, the count of water molecules in the cavity for both DFG-in and DFG-out conformations was performed using the VMD program[175]. As for the SAS analysis, the calculation was repeated 100 times and the number of waters averaged afterwards, determining the water molecules within 5 Å from the residues of the DFG motif.

pK_a calculation

As discussed in the Introduction 1.4.1, it was proposed that the DFG flip transition could be influenced by the protonation of residue Asp404 of the DFG motif. To address this hypothesis, pK_a of Asp404 was calculated for both Src and Abl on 10000 structures with the program PROPKA[176, 177]. Subsequently, the values were averaged.

The structures were extracted from the short MD simulations (20 ns) performed in the DFG-in and DFG-out conformations (see Section 2.1.5). The program PROPKA[176, 177] calculates the pK_a value of residue Asp404. It does so, by summing to the intrinsic pK_a of the residue, retrieved by solving the linearised Poisson-Boltzmann equation, the environmental perturbation acting on the residue. The final pK_a perturbation (ΔpK_a) is given by the sum of the hydrogen bonds contribution, charge-charge interactions and desolvation effect depending on the protein burial.

Allosteric coupling analysis

Allosteric coupling analysis was performed in order to identify correlated motions along normal modes in TKs. Normal modes is a fast method to reconstruct patterns of motion in proteins, calculating vibrational modes and protein flexibility. The method of Balabin *et al.* was applied[178]. It consists of a combination of normal modes analysis and first order perturbation theory.

The normal modes of Src, Abl, Kit, Met and EGFR were retrieved using the Anisotropic Network Model[178, 179], taking into consideration only the Cα atoms of the minimised structure. Subsequently, a first order perturbation is applied to every Cα atom, treated as a spring and displaced along the normal mode[180]. Then the alignment between the displaced (Figure 2.3, coloured in magenta) and the starting structures (see Figure 2.3, in blue) is performed to eliminate translational and rotational motions. In this way, it is possible to identify just the structural perturbation and better distinguish between correlated and non correlated regions. Finally, the RMSD between the aligned (see Figure 2.3, in red) and the starting structures gives the local perturbation for each Cα atom (see Figure 2.3), considering 10 Å cut-off for the neighbour list. The net allosteric coupling is measured on the total displacement of every Cα atom along all the normal modes.

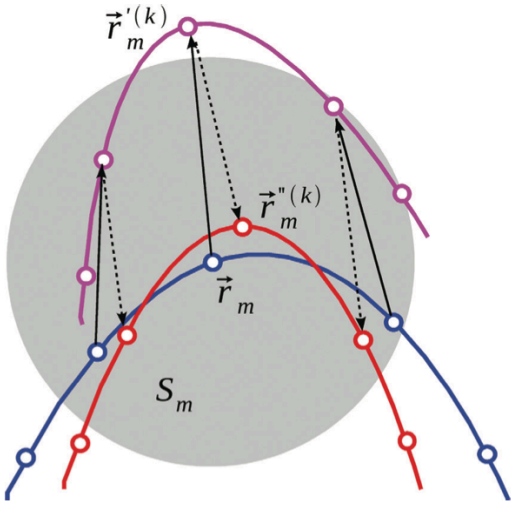

Figure 2.3: Graphical representation of the coarse-grained atom displacement as described in the method of Balabin *et al.*[178]. Figure taken from Ref [178].

2.1.6 Enhanced sampling simulations details

Well-tempering PTmetaD

Well-tempered PTmetaD method[119, 151] was used in the work presented in this thesis to accelerate the sampling of large conformational changes and to reconstruct an accurate and reliable free-energy for Src and Abl. The choice of the computational expensive PTmetaD method is due to the complexity of the biological transition of interest for this study, the DFG-flip (DFG-in to DFG-out).

Simulations of Src and Abl KD were performed using 28 well-tempered PTmetaD replicas at different temperatures, ranging from 308K to 399K. The topology files were prepared using AMBER[133], as described in Section 2.1.4, and converted using Acpype[181] into GROMACS[134] input files. The systems were minimised with 1000 steps of steepest descent and subsequently equilibrated. A first equilibration step was performed in the NPT ensemble for 10 ns, using a Parrinello-Rahman barostat[143], followed by a second equilibration step of 10 ns canonical ensemble, at 300K with a velocity-rescale thermostat (described by Bussi *et al.* in Ref[142]).

For the FE simulations, version 4.6 of the molecular dynamics program GROMACS[134] was used, with the PLUMED plug-in (version 1.3)[163], running on 280 Xeon Nehalem cores connected by an Infiniband network. The van der Waals interactions were cut off at 1.1 nm, and the long-range electrostatic interactions calcu-

lated by the PME algorithm[139] with a mesh spaced of 0.12 nm. The neighbour list for the non-bonded interactions was updated every 0.02 ps. The systems were sampled in the NVT ensemble, coupled with a velocity-rescale thermostat[142] and a time step of 2 fs.

An exchange of coordinates between replicas at two adjacent temperatures is attempted every 1 ps. The initial height of the Gaussians was 2 kJ/mol, the bias factor 5 and the deposition rate of 1/2000 steps. The width of the Gaussians, that determines the resolution of the recovered free energy surface, was set to 0.03 nm for the distance CVs (described below) and to 0.1 nm for the dihedral combination CV. Every well-tempered PTmetaD run required almost 3 months of uninterrupted simulations on 280 processor cores, for a total of 400 ns performed for each of the 28 replicas.

As mentioned previously (Section 2.1.3), in the PTmetaD method is necessary the identification of a set of CVs able to describe the process of interest. In this thesis, to identify the CVs that better describe and discriminate conformational change of the DFG motif, crystal structures of Src and Abl were analysed, together with the trajectories of the MD simulations performed.

Four CVs able to give a geometrical description of the flip were identified. The set used in the PTmetaD simulations of Src and Abl comprises three distances and one dihedral combination, described as follows (see Figure 2.4):

- CV1 is the distance between residue <D404> of the DFG motif and residue <K295> of the β3 strand (<> denotes the centre of mass of the residue). CV1 was aimed to describe the flip of D404 thus, as counterpart residue K295 having a limited range of movements was chosen (see Figure 2.4).

- CV2 is the distance between residue <F405> of the DFG motif and L317(C_β) for Src/I293(C_β) for Abl, residues of the αC-helix (see Figure 2.4). Residues L317 and I293 were chosen due to a peculiar movement of residue I293 of Abl, detected analysing the MD trajectory.

- CV3 is a dihedral combination of residues D404, F405 and A408, $f(\phi 404, \psi 405, \psi 408)$ as shown in Figure 2.4. f is a continuous function ranging from 3, if the three dihedral arguments correspond to the DFG-in position, to 0 if they are in the DFG-out conformation.

- CV4 is the distance between residues <N381,...,H384> and residues <A408,...,I411> (see Figure 2.4). CV4 was aimed to describe the movement of the N-terminal part of the A-loop, region immediately after the DFG motif.

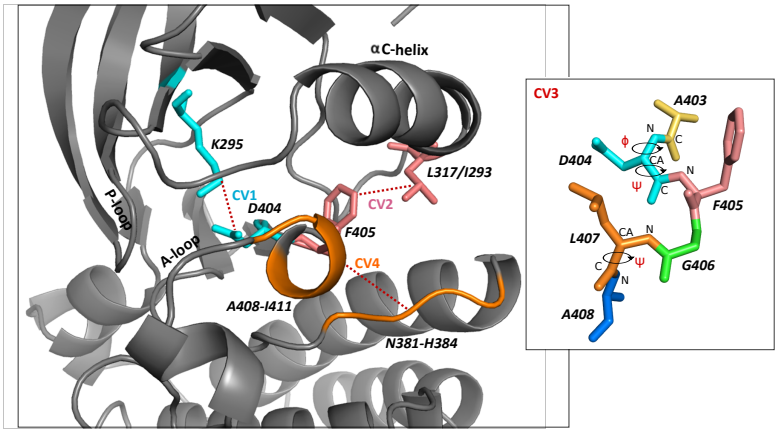

Figure 2.4: Collective variables chosen for the well-tempered PTmetaD simulations of Src and Abl. 2SRC.pdb numbering considered, with the exception of the residue I293 in Abl that corresponds to L317 in Src. In detail, the CV3 dihedral combination of the DFG motif. Figure taken from Ref [93].

PTmetaD in well-tempered ensemble

The well-tempered PTmetaD method[119, 151] allows the accurate sampling of rare events in biological systems. However, the high computational cost depending on the number of replicas needed to simulate large systems (e.g. more than 35000 atoms) is still a huge limitation. The main advantage in using the PTmetaD in a well-tempered ensemble (WTE)[182, 183], is the possibility to obtain good performances (high exchanging rates among temperatures), reducing the number of replicas. In the WTE ensemble, the average potential energy remains approximately the same of the canonical ensemble, but with enhanced fluctuations (see Figure 2.5). The enhancement of the fluctuations is obtained by tuning the γ factor, to ensure a better overlap between the potential energy of each replica:

$$V(U,t\to\infty) = -(1-\gamma^{-1})F(U) \tag{2.6}$$

with γ=(T+ΔT)/T $\geq$ 1 and F(U)=U-β^{-1} lnN(U)).

In practice, the first 20 ns of simulation were used to converge the bias in the only potential energy space. Subsequently, PTmetaD-WTE simulations were continued

considering the other CVs, adding the history-dependent metaD bias to the static bias, previously generated. At the end of the simulation, both the bias and the trajectory corresponding to the first 20 ns were discarded and the FES reconstructed as usual in function of the CVs.

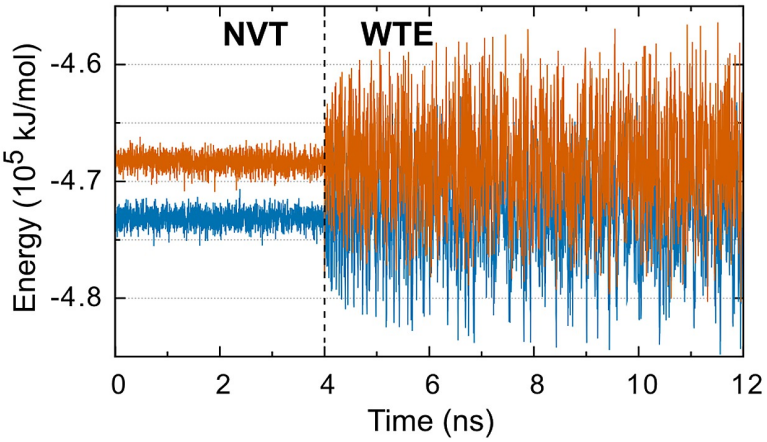

Figure 2.5: Comparison of the potential energy fluctuations for two replicas in a normal NVT simulation and in a WTE simulation. Figure taken from Ref [184].

The simulations of the resistant mutants of Abl (G250E, E279K, H396P, E450K and T315I) were performed using only 5 replicas (compared to the 28 of the well-tempered PTmetaD), using temperatures at 300K, 318K, 337K, 356K and 399K. The systems were simulated as described in Section 2.1.6 using the same set of CVs (see Figure 2.4). The Gaussians added along the potential energy space had initial height 2 kJ/mol and width 140. The bias factor was 16 and the γ factor was 2.24 (square root of the bias factor). The simulations converged after reaching 400 ns with the exception of G250E and T315I, converged at 640 ns and 1200 ns, respectively. Overall, the sampling time was of 15 μs across all mutants.

Metadynamics using path collective variables

The path collective variables (PCV)[154], implemented in PLUMED[163], were used as CVs in the Metadynamics simulations to reconstruct the free energy of the drug imatinib binding to Abl WT and the "gatekeeper" mutant (T315I). The starting structure used was the crystal structure of Abl with imatinib bound (PDB code 1IEP). The mutant T315I with imatinib bound was modelled in silico using Modeller[167].

The path collective variables[154] are meant to find low free energy pathways along the system reaction coordinate. Two CVs, the $s-path$ and the $z-path$, are used

to describe the evolution of the system along a predefined path. The path is constituted by a set of conformations (X) that represent all the relevant intermediates describing the transition from X_A to X_B.

The $s-path$ describes the progress of the system along the path and is defined as follows:

$$s-path = \frac{\sum_{i=1}^{N} i \exp(-\lambda d[X-X_i]^2)}{\sum_{i=1}^{N} \exp(-\lambda d[X-X_i]^2)} \tag{2.7}$$

The $z-path$ describes the distance of the system from the path:

$$z-path = -\frac{1}{\lambda} \ln \left[\sum_{i=1}^{N} \exp(-\lambda d[X-X_i]^2) \right] \tag{2.8}$$

where N is the number of conformations X_i that constitute the path. At every time step is calculated the RMSD distance of the system from the path, $d[X-X_i]^2$.

For the path to be meaningful, the conformations should be chosen as equidistant as possible, in a way that the system could follow the path without discontinuities. To obtain a smooth path the term λ should be chosen as follows:

$$\lambda = \frac{2.3(N-1)}{\sum_{i=1}^{N-1} d[X_i - X_{i+1}]^2} \tag{2.9}$$

For this thesis, a suitable path describing the (un-)binding mechanism of imatinib to Abl was extracted from a preliminary Metadynamics simulation, performed by Mr Michael Hayne a previous student in the group. Sixteen conformations (considering the only Cα atoms) were extracted from the lowest free-energy pathway of the metaD simulation, and then used as CV ($s-path$). The $z-path$, distance from the path, was set by defining an upper wall (UWALL=1) in the PLUMED input file. The walls are constraints usually used to limit the CV space explored. The same path ($s-path$) was used for the simulations of the T315I mutant.

2.1.7 Rational choice of PTmetaD parameters

As mentioned previously in Section 2.1.3, when performing a PTmetaD simulation some parameters have to be chosen by the user, and this choice could determine the outcome of the simulation. In particular, speed and accuracy may depend on the height

(w), width (σ) and deposition rate (τ_G) of the Gaussians. Hereafter, I briefly describe the best way to choose these and other parameters, in order to assure a reliable sampling. Figure 2.6 shows an example of the PLUMED[163] input file needed to perform PTmetaD simulations.

If the height (HEIGHT in Figure 2.6) and width (SIGMA in Figure 2.6) of the Gaussians (HILLS in Figure 2.6) are too high, the resolution of the free-energy reconstruction will be compromised while, if they are too low the simulation will take very long to fill the energy minima and to converge. Thus, as a rule of thumb:

- the W_STRIDE, in Figure 2.6, defines the deposition rate of the Gaussians and how often (time steps) is written the HILLS output file. It is usually chosen to be 1 ps.

- the HEIGHT of the Gaussians should be usually around 1.5-2 kJ/mol. It could be higher or lower, depending on the system and how deep is expected to be the energy minimum.

- the SIGMA should correspond to half of the CV fluctuation at equilibrium (expressed in nm). This is usually determined on the basis of preliminary MD simulations.

- the BIASFACTOR (Figure 2.6) is chosen between 5 and 10, depending on the expected height of the energy barrier to cross.

In the input file, the CVs to consider during the simulation are defined (CV DEFINITION in Figure 2.6) by specifying the atoms or groups of atoms constituting each of the CV (GROUP DEFINITIONS in Figure 2.6). The values of the CVs along the simulation are written in the COLVAR output file, with a frequency specified by the command PRINT W_STRIDE (see Figure 2.6). Additional parameters are the upper and lower walls (UWALL, LWALL in Figure 2.6), constraints used to limit the CV space to explore. If the CV value becomes greater than the fixed wall limit (LIMIT in Figure 2.6) then, a restraining potential starts acting on the system, with KAPPA (Figure 2.6) being the energy constant of the potential.

```
HILLS HEIGHT 2.0 W_STRIDE 1000  Gaussians
PRINT W_STRIDE 500 how often is saved the COLVAR file

PTMETAD  method of choice

WELLTEMPERED SIMTEMP 310 BIASFACTOR 5.0 simulation temperature

##########################################
# CV DEFINITION
##########################################
#CV1 :
DISTANCE LIST <lys40> <asp150> SIGMA 0.03 residue center of mass
UWALL CV 1 LIMIT 3.50 KAPPA 100.0

#CV2 : ring PHE151 e CB ILE62
DISTANCE LIST <phe151ring> 1006 SIGMA 0.03
UWALL CV 3 LIMIT 3.50 KAPPA 100.0

#CV3 : DFG dihedral combination
#phi ALA149: C ASP150: N CA C
#psi ASP150: N CA C PHE151: N
#psi LEU153: N CA C SER154: N
ALPHABETA NDIH 3 SIGMA 0.1
2438 2440 2442 2450 1.0
2440 2442 2450 2452 0.5
2479 2481 2496 2498 -0.5

#CV4 :
DISTANCE LIST <beta127_130> <beta154_157> SIGMA 0.03
UWALL CV 2 LIMIT 4.00 KAPPA 100.0

##########################################
# GROUP DEFINITIONS
##########################################
lys40->           atoms defining the center of mass
647 650 653 656   of every residue or group of residues
lys40<-

asp150->
2447 2448 2449
asp150<-

beta127_130->
2077 2091 2111 2130
beta127_130<-

beta154_157->
2500 2511 2535 2556
beta154_157<-

phe151ring->
2459 2460 2462 2464 2466 2468
phe151ring<-

##########################################
ALIGN_ATOMS LIST <allused>
allused->
LOOP 640 2600 1
allused<-
##########################################
```

Figure 2.6: Example of the PLUMED input file (version 1.3[163]) used to perform the PT-metaD simulations described in this thesis. Highlighted in yellow are the parameters chosen by the user. The numbering used for the residues defining the CVs do not correspond to the 2SRC.pdb numbering, used in Section 2.1.6, but to the original pdb file used in the simulation.

2.1.8 PTmetaD analysis

Free energy surface reconstruction

The free energy surface (FES) of the PTmetaD simulations was reconstructed as a function of the biased CVs. The deposited energy bias along the trajectory was integrated for the replica at 310K. The HILLS output file was given as input to the sum_hills post-processing program, implemented in the PLUMED plugin[163]. In the HILLS file are stored all the Gaussians deposited during the simulation for each CV. By specifying in the command line of sum_hills the CVs to consider, the program integrates the added bias over time, and it reconstructs the free energy as function of the selected CVs.

Checking the convergence of the free energy reconstruction

The convergence of free energy (FE) calculations is a serious concern of enhanced sampling methods. In Metadynamics usually only when the system is able to diffuse freely in the entire CV space and the variations in FE are minimum, the FE reconstruction is converged. The convergence of the FE can be monitored by reconstructing the added bias (sum_hills) at regular intervals. If the FE variation is minimal (less than 1 kcal/mol) during the last 50 ns of simulation, the calculation is usually converged.

Mono-dimensional and bi-dimensional reconstruction of the FE, for the simulations performed in this thesis, were obtained and checked for convergence. Only projections along CV1 (distance D404-K295) and CV2 (distance F405-L317/I293) are shown throughout this thesis (see Results), because they are the two CVs that better discriminate between the DFG-in and DFG-out conformation.

Free energy reweighting

The reweighting technique[185, 186] (performed with driver, the post-processing program of PLUMED[163]) was used to monitor CVs not directly sampled during the PTmetaD simulation. This is possible thanks to the enhanced sampling guaranteed by the Parallel Tempering method[157]. All conformational changes happening on a faster timescale than those described by the CVs, will be sampled as well during the simulation, making their reweight possible later on. Thus, additional CVs could be chosen *a posteriori* and their FE profile retrieved by reweighting the added bias.

The PTmetaD reweighting performed in the study was meant to investigate the N-lobe dynamics in Src and Abl. The FE profiles were reweighted along four new CVs

that describe the movements of the N-lobe loops and αC-helix. The selected distances were (see Figure 2.7):

- CV5 = distance between residues <F278> of the P-loop and <P304> at the N-terminal of the αC-helix
- CV6 = distance between residues <P304> of the αC-helix and <E331> of the β2-β3 loop
- CV7 = distance between residues <K316> at the C-terminal of the αC-helix and <M380> at the N-terminal of the αF-helix
- CV8 = distance between residues <K298> and <M302> both belonging to the β3-αC loop

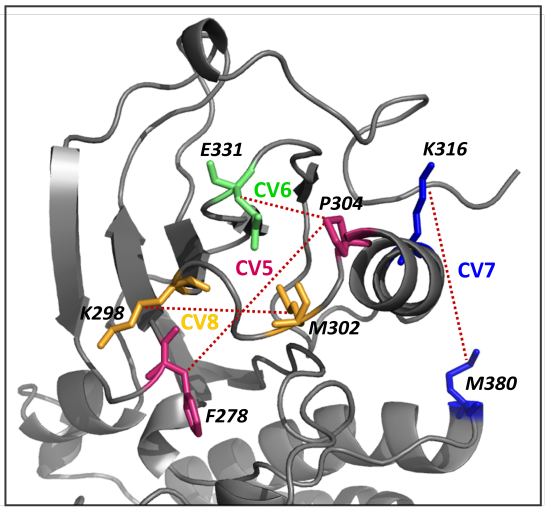

Figure 2.7: Collective variables chosen for the reweighting of the well-tempered PTmetaD simulation for both Src and Abl (2SRC.pdb numbering considered). Figure taken from Ref [187].

The second PTmetaD reweighting performed interests the resistant mutants of Abl and the wild-type Abl and Src. It was meant to investigate the conformational change of the activation loop, and in particular the population of the open conformation sampled by the mutants. The FE profiles were reweighted along two path collective variables describing the "open to close" path of the A-loop (Figure 2.8) and having as reference conformations the crystal structures of Src and Abl with A-loop open and closed (2SRC.pdb (closed) and 1Y57.pdb (open) for Src; 1IEP.pdb (closed) and 2F4J.pdb

(open) for Abl). CV1 corresponds to the A-loop path "open to close" in Src and CV2 to the A-loop path "open to close" in Abl.

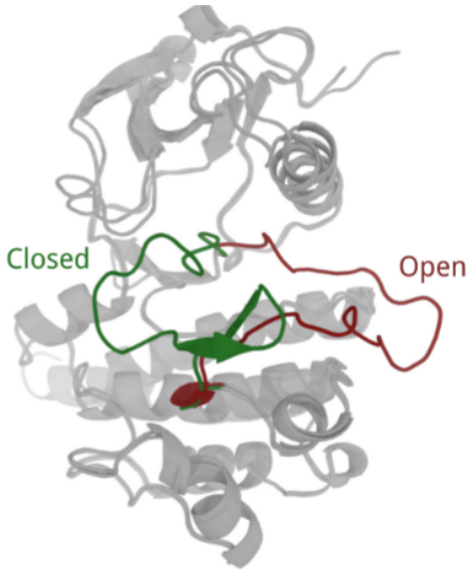

Figure 2.8: Representation of the activation loop in the open (in red) and closed (in green) conformations used as references in the PTmetaD reweighting of Src and Abl.

Structural analysis of the FES

The structural analysis of the PTmetaD simulations was performed extracting snapshots from the trajectory corresponding to the FE minima and transition states of the free energy surface (FES).

2.1.9 Limitations of the computational methods

As I have previuosly discussed, the main factors limiting the accuracy of MD simulations are: the quality of the force-field used and the timescale problem[188]. Protein force-fields trade accuracy for speed. The hydrogen bonds, for example, are described as Coulomb interactions despite their quantum mechanical nature[189]. Moreover the electrostatic and the van der Waals interactions are approximated. Polarizability is not included and the Lennard-Jones potential is used to define dispersion forces[190]. However, the significant efforts made in the last years to include the missing terms in a mean field average manner, make protein force-fields more accurate and reliable[191].

In a recent paper, Lindorff-Larsen *et al.*[132] compared the performance of several recent force-fields in protein folding (see Figure2.9). The force-fields with the lower score, and so with the higher agreement with experiments, are Amberff99SB*-ILDN[128, 168] (applied in this thesis) and CHARMM22*[131].

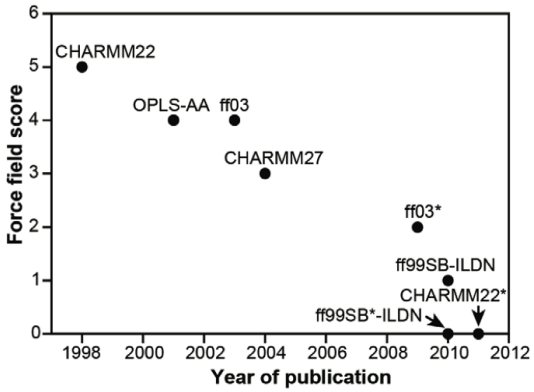

Figure 2.9: Improvement of force-fields accuracy over time. The force-fields with lower scores are those showing a higher agreement with experiments. Figure taken from Ref [132].

With respect to the timescale problem, I have discussed the pros and cons of enhanced-sampling methods. As nowadays it is still impractical to simulate an entire protein domain for milliseconds, even using specialised hardware[118], in order to study complex transitions the use of enhanced-sampling approaches is necessary[145, 144].

2.2 Experimental Methods

For the purpose of this study, molecular biology and biophysical experiments were performed to investigate the affinity of the drug imatinib towards engineered mutants of the Src kinase. Mutations have been engineered on the basis of the computational data retrieved in the work. The *in vitro* expression of Src and its mutants was set up to perform isothermal titration calorimetry experiments (ITC).

2.2.1 *In vitro* protein expression

Traditionally, the expression of recombinant proteins is achieved by infecting cells with a DNA vector, containing the template for the expression of the protein of interest. The infected cells are then cultured and the transcription of the desired protein activated. For these purposes, *E.coli* bacterial cells are widely used, because they are easy to culture and able to assure high protein yields[192]. Subsequently, the cells expressing the recombinant protein are lysed and the protein extract is purified. The selection of cell types (*e.g.* bacterial, mammalian, insect), growing medium (*e.g.* Luria Broth medium, autoinduction medium, minimum medium) and growing conditions (temperature) depend on the desired yield and type of recombinant protein to express[192].

In the present work *E.coli* strain BL21(DE3) cells were used. The plasmid encoded for the unphosphorylated form of Src kinase domain (supplemented with an hexahistidine tag followed by a TEV protease cleavage site used for purification purposes), plus for the YopH phosphatase and the chaperone pG-KJE8 (Takara), encoding for groESL proteins. The co-expression of Src KD together with phosphatase[20] and chaperones[193] was meant to obtain unphosphorylated and properly folded protein, respectively[194]. The same construct was used also for the expression of the resistant mutants of Src.

2.2.2 Protein purification

Chromatography is one of the most widely used means to achieve the separation of individual proteins from protein mixtures. The choice of the chromatography techniques to adopt in the design of a purification protocol may vary. The choice is usually based on the physical and chemical properties of the recombinant protein to isolate. The purification protocol is carefully tailored for every protein. In fact, what could be successfully used to purify one protein, might determine the inactivation of another[195, 196, 197, 198].

In practical terms, the protein extract (also called surnatant), obtained after cells lysis, is loaded onto the chromatography column. General properties of proteins, such as charge, size or specific interactions with engineered protein tags, are exploited in the separation process. The chromatography columns are composed of a packed matrix (*e.g.* agarose, silica, dextran), called the stationary phase. Loaded onto the column and moving throughout it are the eluents, *i.e.* the mobile phase, such as the protein mixtures and the buffers.

Different chromatography techniques are available[195, 196, 198]. A common mechanism, used in affinity and ion-exchange chromatography, is characterised by a first selective capturing step, where the protein of interest binds to the stationary phase, while impurities are washed away by the eluent. Subsequently, the conditions favouring the binding are changed, so that the protein is finally eluted[195]. The mechanism just described is illustrated in Figure 2.10, example of a chromatogram profile generated during a purification process. However, an exception to this method is size exclusion or gel filtration chromatography, where the protein of interest is separated based on its

size (molecular weight)[195, 196].

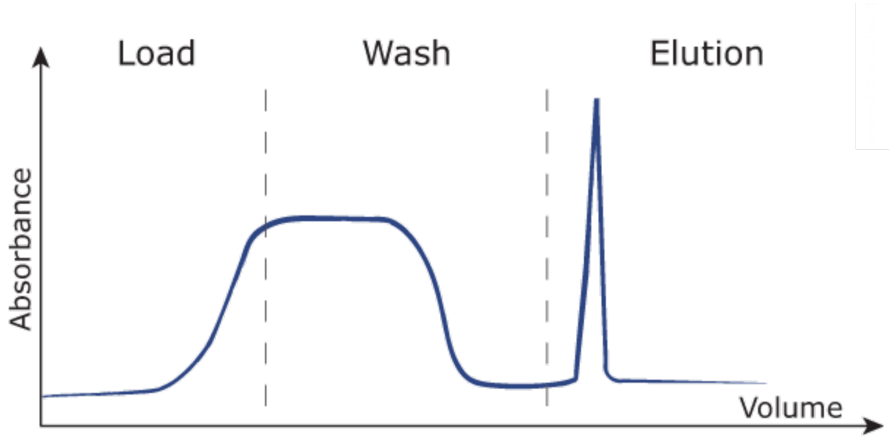

Figure 2.10: Example of the chromatogram generated by affinity and ion-exchange chromatography method. During purification the solution, before being collected, passes through an UV detector. Thus, the UV absorbance profile is monitored to check for protein concentration. After loading the sample into the column (load phase, no UV signal), the protein of interest will bind to the column, while all the impurities will be eluted, determining the increase of the UV signal. It follows a wash phase to eliminate all impurities, and only when the UV profile has returned to baseline, the conditions will be modified to determine the elution of the target protein (elution phase, UV peak). Figure taken from Ref [199].

Affinity chromatography

Affinity chromatography exploits the formation of hydrogen bonds, electrostatic and van der Waals interactions, which are involved in many binding processes, for the separation of proteins[200, 196, 198].

The matrix of the chromatographic column is supplemented with a specific ligand, covalently bound to the matrix and proved to have high affinity for the target protein. When the sample containing the protein mixture is loaded onto the column, the protein of interest binds, reversibly, to the ligand. Thus, while the other proteins will be washed away by the continuos elution of buffer, the target protein will remain stuck to the column. The subsequent change in the conditions of the loaded buffer will weaken the protein-ligand binding thus, causing protein elution.

In this study the chromatographic column used was the HisTrap HP nickel column (GE Healthcare). The column was supplemented with nickel (the ligand) while, as mentioned in Section 2.2.1, a hexahistidine tag (His-tag) was engineeringly attached to the kinase domain of Src and of the mutants. The His-tag has a high affinity for nickel and thus Src will remain initially stuck to the column. Only when a imidazole rich buffer is loaded, Src dissociates and is eluted (histidine has higher affinity for

imidazole than for nickel).

Ion-exchange chromatography

Ion-exchange chromatography bases its separation strength on differences in protein surface charges. The protein of interest could interact with a positively or negatively charged stationary phase, depending on the protein surface charge (see Figure 2.11). The elution of the protein is mediated by changes in the pH or in the ionic strength of the buffer[195, 196, 198].

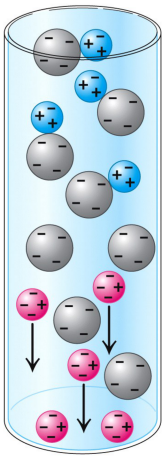

Figure 2.11: Ion-exchange chromatography. The matrix is negatively charged (grey beads). Positively charged proteins (blue beads) interacts with the matrix and are retained, while negatively charged proteins (pink beads) are eluted. Figure taken from Ref [201].

In this study the Ion-exchange chromatographic column used was the Q anion exchange column (GE Healthcare). The elution of the target protein was obtained by changing the ionic strength of the buffer. Once the protein was stuck to the column, a buffer with high concentration of NaCl was loaded. These ions compete with the protein to interact with the matrix and thus, the stronger is the protein charge the higher should be the salt concentration needed.

In ion-exchange chromatography, as well as in affinity chromatography, buffer changes were performed by gradient elution. Let's consider buffer A as the buffer favouring the binding of the protein to the column, and buffer B the one disfavouring the binding. The gradient guarantees not to have a sudden change between buffers, but to have a progressive increase of buffer B over buffer A, till reaching 100% of buffer B at the end of the gradient.

Size-exclusion or gel filtration chromatography

As previously mentioned (see Section 2.2.2), size-exclusion chromatography separates proteins according to their size. The column matrix does not have any bound ligand, but it is just constituted of porous particles[195, 196, 198]. The porous dimension could be chosen accordingly to the molecular weight of the target protein, in order to obtain a good separation among the protein mixture. The column used in the study was a Superdex 75 16/60 size-exclusion column (GE Healthcare).

Proteins, based on their size, can follow an easier or more difficult pathway throughout the column. Bigger size proteins, for steric reasons, will be eluted first because they are excluded from the column matrix, while smaller proteins will be retained longer in the intricate maze of the column matrix, delaying their elution[195] (see Figure 2.12).

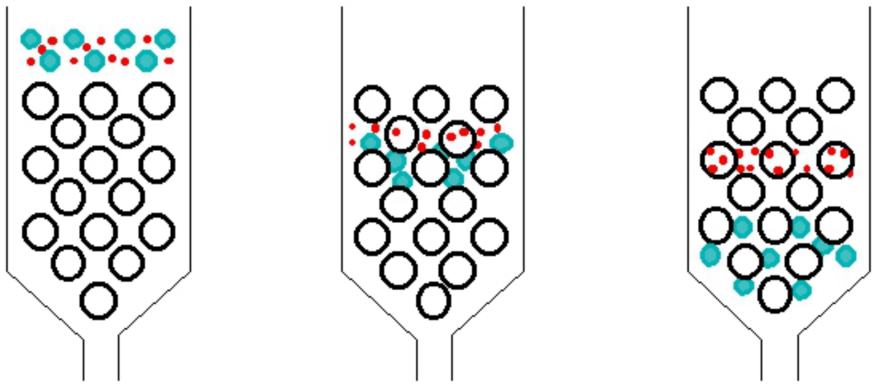

Figure 2.12: Size-exclusion chromatography. Large proteins are excluded by the column beads, while small proteins are retained and eluted later. Figure taken from Ref [202].

2.2.3 Expression and purification of Src wild-type and Src engineered mutants

Expression of Src wild-type for isothermal titration calorimetry

Src KD was expressed and purified following the procedure published in Ref [194] with two exceptions regarding the labelling of the protein and the affinity tag cleavage, both avoided since not crucial for ITC experiments (see Figure 2.13).

The *E.coli* cells were grown at 37°C in 1 litre of Luria Broth standard medium, containing the specified amounts of antibiotics: 50 μg/ml of kanamicin, 50 μg/ml of streptomycin and 20 μg/ml of chloramphenicol, for Src KD and YopH plasmid se-

lection and 5 ng/ml of tetracyclin, for continuous induction of chaperones expression. Reached the optical density (OD_{600}) of $\sim$0.8, protein expression was induced by adding to the cell culture 0.2 mM of IPTG, the temperature was lowered at 20°C and cells were left growing overnight.

Expression of Src engineered mutants for isothermal titration calorimetry

The same expression protocol of Src WT was successfully used for the production of the engineered αG mutant (Q275A/P299Q/V461S) of Src. Due to protein solubility problems, a new protocol (see Figure 2.13) was established for the expression of the other mutants (Q275A/P299Q/Q420E, Q275G/P299E/Q420A and αC-chimera). In the new protocol, cells were grown at 37°C in 1 litre ZYP-5052 autoinduction medium[203], supplied with antibiotics (for quantities see previous Section). At OD_{600} $\sim$0.8, temperature was lowered at 16°C and expression was allowed for 48 h, to reach a good bacterial expression. The adoption of the autoinduction medium together with the change in temperature (from 20°C to 16°C), were found to be crucial for a proper expression of the Src mutants, because the slow cell growth assured the correct folding of the protein, avoiding precipitation.

Purification protocol for isothermal titration calorimetry

The cell culture expressing the protein was centrifuged at 4000 rpm for 15 minutes at 4°C. The pellet was resuspended with buffer (50 mM Tris pH 8, 500 mM NaCl, 25 mM Imidazole and 5% Glycerol) and it was sonicated on ice (instrument settings: 37% power, 3 seconds on and 1 second off pulser for a total of 10 minutes). The sonicated sample was then centrifuged at 40000 rpm for 40 minutes at 4°C. The surnatant, containing the protein of interest, is now ready for purification.

A new two-step purification protocol was developed in the work presented in this thesis (see Figure 2.13). It allows one to obtain greater yields of high quality proteins (needed for ITC experiments) and to reduce the experimental time needed (1 day long purification). After sonication, the surnatant was loaded on the HisTrap HP nickel column (GE Healthcare) and the protein was eluted with a linear gradient of 0-100% of elution buffer (50 mM Tris pH 8, 500 mM NaCl, 500 mM Imidazole and 5% Glycerol). This first purification step is a rough separation made with a nickel column, through affinity chromatography. The protein construct was supplied of an His-tag

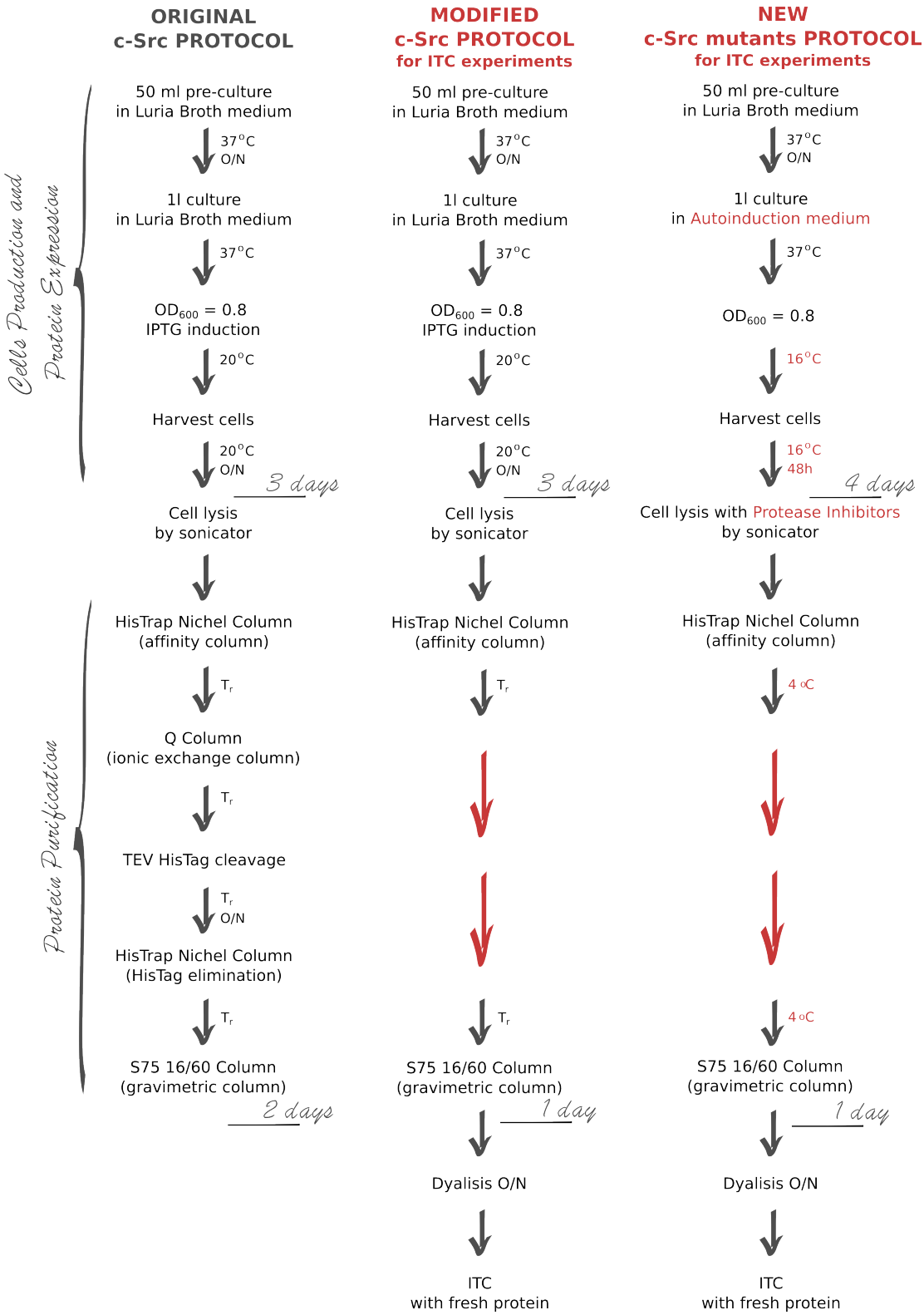

Figure 2.13: Schematic workflow of the protocols used for the expression and purification of Src WT and the engineered mutants of Src. In red are highlighted the changes made to the original protocol of Ref [194].

(hexahistidine tail) with an high affinity for nickel but even higher for imidazole. When the protein passes through the HisTrap column, it remains stuck and only when the imidazole rich buffer is loaded, the protein can be eluted.

The recollected fractions from the HisTrap column with the target protein were then loaded onto a Superdex 75 16/60 size-exclusion column (GE Healthcare), equilibrated with a buffer composed of 20 mM Tris pH 8, 0.5 mM NaCl, 1 mM $MgCl_2$ and 1 mM TCEP. The second purification step permits to obtain a pure product, since separation of proteins on the size-exclusion column (S75) happens by means of their molecular weights.

The simplified protocol skips two of the four steps of Ref [194] (see Figure 2.13): the Q column and the TEV-cleavage. Both were avoided because they cause a huge loss in protein yields. Moreover, the TEV cleavage was not needed, since the presence of the His-Tag is not detrimental for ITC experiments. A further implementation to the protocol was necessary given the high instability shown by the engineered mutants of Src at room temperature (T_r). Thus, the entire purification was carried out in a cold room at 4°C.

2.2.4 Isothermal titration calorimetry

The study of the binding between a ligand and its substrate could be done by means of different experimental techniques: fluorescence assays, NMR, surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC). ITC is one of the most widely used techniques in drug discovery[204]. The absence of probes (as in fluorescence), of temperature, molecular weight and pH limitations (as in NMR) and neither the need to immobilise the substrate (as in SPR), are all advantages rendering ITC assays very straightforward and reduced in costs. However, the large amount of sample needed and the low sensitivity of the instrument for nM binding range, are the reasons why sometimes ITC is not the best choice. Nevertheless, the advantage of using this technique is the possibility to rapidly recover the thermodynamics of the binding process of interest[205, 206]. In this study, ITC experiments were performed to determine the affinity of imatinib for Src WT kinase and for the engineered mutants of Src.

The instrument is constituted of two cells (see Figure 2.14). The sample cell, that is filled with the substrate in solution (Src), and the reference cell, filled with water

and maintained at a constant temperature. During the experiment, the solution of the ligand (imatinib) is titrated with a syringe into the sample cell. The binding event of the ligand with the substrate will produce, or not, heat (exothermic or endothermic reaction). Changes in temperatures between the cells are detected by the instrument. The quantification of the heat produced by the binding events, derives by the power needed to maintain the two cells at the same temperature (ΔT=0).

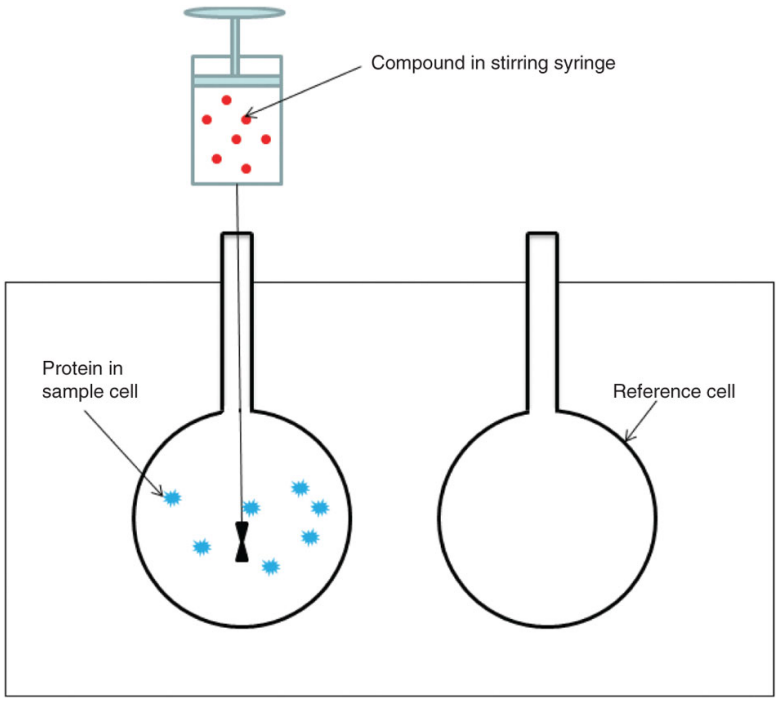

Figure 2.14: Schematic representation of an ITC instrument. Figure taken from Ref [207].

A graphical representation of the titration is created in real time, with peaks referring to the variation in power (μcal) over time. The area under the peaks is then integrated to recover the enthalpy corresponding to the binding event (see Figure 2.15). The heat (kcal/mol) derived from the formation of the protein-imatinib complex, in function of the molar ratio of the ligand, produces the isothermal curve, whose slope corresponds to the dissociation constant, K_d (see Figure 2.15). The ITC experiment permits to directly retrieve the binding affinity ($K_a = 1/K_d$), the enthalpy (ΔH) and the stoichiometry (n) of the binding event under analysis. Moreover, given the relationship:

$$\Delta G = -RT \ln K_a = \Delta H - T\Delta S \tag{2.10}$$

the variation in the free energy of binding (ΔG) and in entropy (ΔS) could be recovered

as well.

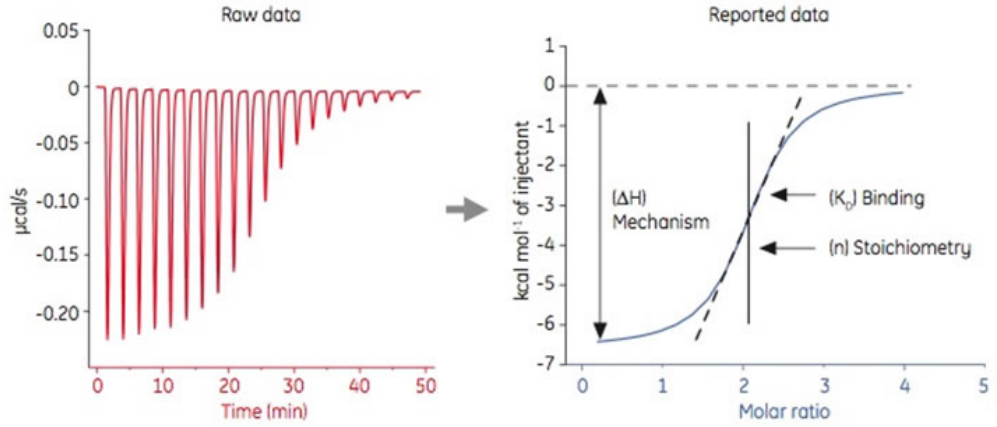

Figure 2.15: Example of the outputs produced during an ITC experiment. Figure taken from Ref [208].

Particularly important for the accuracy of the results is for the two solutions, the one with the substrate and the one with the ligand, to be identical. The thermodynamic parameters, for the binding of imatinib to Src WT and to the engineered mutants of Src, were acquired using a VP-ITC microcalorimeter (MicroCal Inc.).

Sample preparation for ITC experiments

The purified protein was dialysed over night in buffer MES with pH 6.5 (50 mM MES, 500 mM NaCl, 1 mM EDTA, 5 % (v/v) Glycerol). Protein concentration was determined by NanoDrop 1000 spectrophotometer (λ=280 nm) and immediately used to perform the experiment.

ITC essay details

The protocol for the ITC experiment of Src is inspired by the work collected in the doctoral thesis of Dr Ralitza Boubeva[209]. Some changes have been introduced to improve the experiment accuracy.

The sample solution consisted of 2 ml 0.1 mM protein in MES buffer (pH 6.5) and supplemented with 2% DMSO. The ligand solution consisted of 2 ml 1.6 mM imatinib in MES buffer, at the same pH and with 2% DMSO. Both solutions were degassed for 5 minutes under vacuum. Titrations were conducted at 30°C and consisted of a first control injection of 2 μl followed by 37 injections, of 8 μl and 16s duration each, spaced by an equilibration period of 480s. The sample during the measurement was stirred at 266 rpm. Raw data were collected and corrected for ligand heats of dilution. A

one site binding model was assumed and data were fit using MicroCal Origin software (version 7.00). All the experiments were repeated in triplicate.

Thank to the 1 day long purification protocol it was possible the use of fresh protein for ITC experiments, assuring an high percentage of protein binding sites availability, increasing the quality and reliability of the essay.

Chapter 3

The different flexibility of Src and Abl kinases regulates the accessibility of the DFG-out conformation

In this Chapter, the two TKs Src and Abl are investigated in great detail, to unveil the reasons underlying the different affinity towards the anti-cancer drug imatinib. As discussed in the Introduction (see Section 1.4.1), Abl is imatinib primary target (IC_{50}=0.2 μM), while Src is weakly inhibited by the drug (IC_{50}>100 μM), despite the 45% sequence identity[65] and the presence of crystal structures showing imatinib binding to both kinases[20].

Src and Abl can adopt the druggable DFG-out conformation, and the corresponding crystal structures show no major differences between their binding interactions with imatinib[78] (see Introduction, Section 1.4.2). It has been proposed that the binding of imatinib follows a conformational selection mechanism[78, 81] (see Section 1.4.2). Therefore, it could be hypothesised that the kinetics and/or thermodynamics of the conformational transitions leading to the DFG flip (conformational change from DFG-in to DFG-out) may determine the affinity to imatinib of a specific kinase.

To test these hypotheses, the DFG flip was studied for both Abl and Src, by means of massive molecular dynamics simulations (MD) and free energy calculations (PT-metaD). Given the complex nature of the conformational change analysed (DFG flip), long classical MD simulations (1 μs long) were not sufficient to obtain a satisfactory

sampling of the DFG transition. Thus, the use of advanced methods (PTmetaD), able to enhance the sampling of slow conformational changes, was essential.

The results reported below point out crucial differences in the accessibility and stability of the DFG-out conformation of Src and Abl, ascribable to structural and dynamical features proper of the Abl kinase domain (KD). These peculiarities in both the sub-μs dynamics and conformational dynamics of Abl could be responsible for the higher imatinib activity reported for Abl.

3.1 Classical MD simulations and RMSF analysis of Src and Abl

Classical 1 μs long MD simulations were performed using as starting structures the DFG-in active conformation of Abl and Src kinase domain (see MD details Section 2.1.4). Classical MD simulations were extensively exploited to study the essential dynamics (sub-μs dynamics and dominant motions) of TKs.

To investigate the sub-μs dynamics (pico to nanosecond timescale) of Src and Abl, the root mean square fluctuation analysis (RMSF, see Materials and Methods, Section 2.1.5) was used. The RMSF analysis performed on the MD trajectories reveals different flexibility patterns in the sub-μs dynamics for the two proteins. Higher average fluctuations were observed for Abl with respect to Src, especially in the N-lobe loops (P-loop, β2-β3 loop and β3-αC loop) and in the A-loop. The only exception was the αG-helix region, more flexible in Src (see Figure 3.1). All the mentioned areas, which show significant differences in RMSF between both systems, are involved in crucial conformational changes linked to the DFG flip[17].

The analysis of the MD trajectories shows a peculiar conformation of the αC-helix region that differs in the two structures. The αC-helix of Abl experiences an upwards movement (see Figure 3.2). This movement could be relevant, not only because the αC-helix is involved in the DFG-in to DFG-out switch, but also because of its vicinity to the binding site, where imatinib docks.

In order to characterise this conformational change in more detail, the distances between the Phe405 of the DFG motif and the αC-helix residue Leu317 in Src and Ile293 in Abl, were monitored (2SRC.pdb numbering used). These two residues

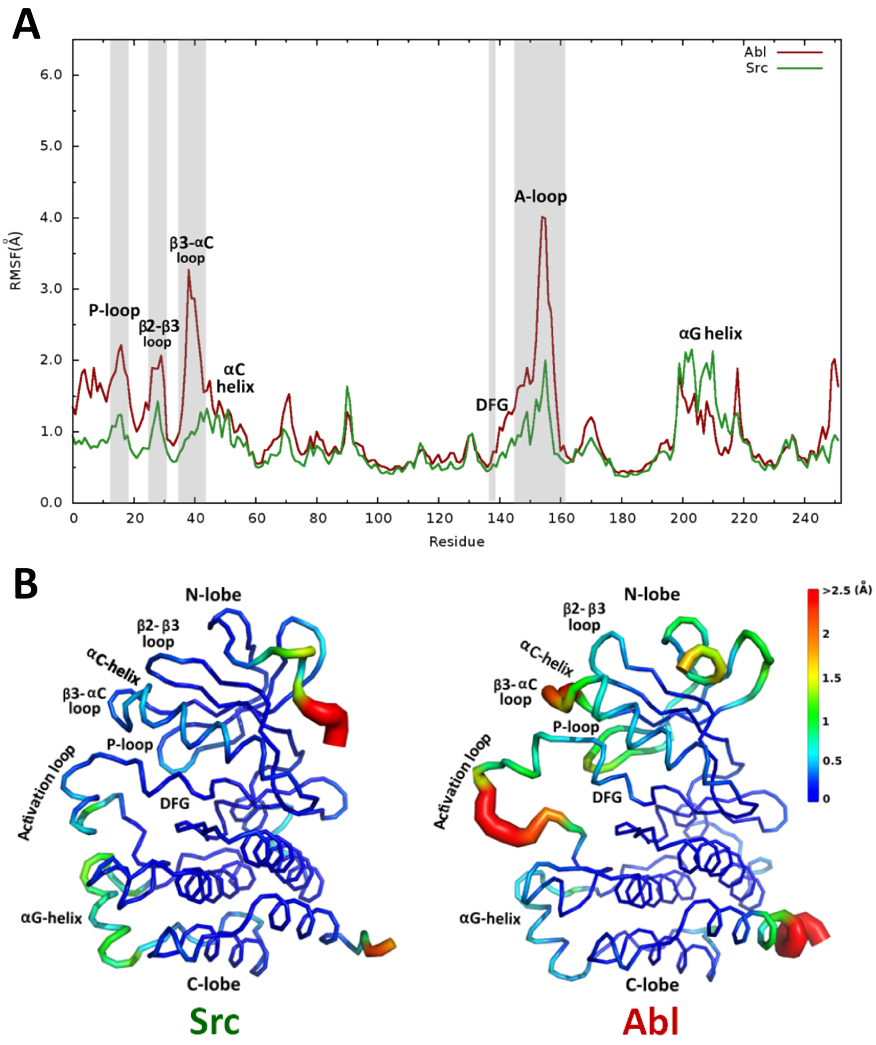

Figure 3.1: (A) RMSF profiles of the C_α atoms of Src (in green) and Abl (in red). (B) RMSF analysis representation directly on the structures of Abl and Src. The thickness of the ribbon is proportional to the RMSF magnitude. The colour spectrum ranges from more rigid (blue) to more flexible (red). Figure taken from Ref [187].

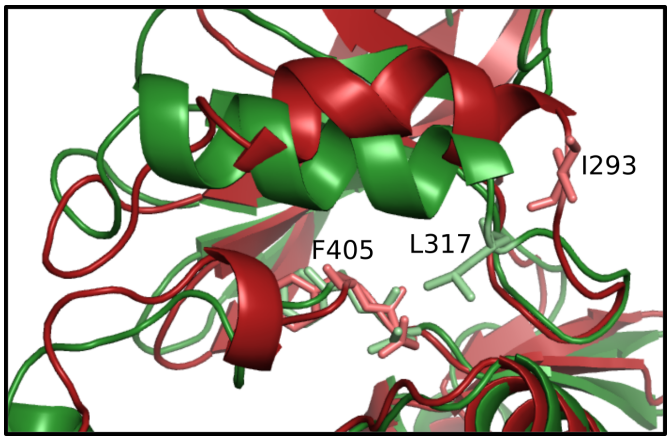

Figure 3.2: Conformations adopted by the αC-helix of Src (in green) and Abl (in red). The distances F405-L317 in Src and F405-I293 in Abl (residues shown in sticks) were monitored to characterise the αC-helix movement.

(Phe405 and Leu317/Ile293) belong to the allosteric pocket of the active site, and are also part of the regulatory hydrophobic spine (R-spine, see Introduction 1.2.2)[23, 22] thus, involved in the regulatory network of the kinase.

In Src, the mean distance between the residues (considering the C_ζ atom of Phe405 and the C_β atom of Leu317) fluctuates around a value of 4.6 $\pm$ 0.4 Å, while in Abl it reaches a larger average of 5.2 $\pm$ 0.6 Å (see Figure 3.3). The significant increase in the distance reached by Abl (with maximum up to 8.5 Å) is due to a side-chain flip of residue Ile293 (see Figure 3.2). From the trajectory, it could be seen that the side chain briefly jumps from its usual "in" conformation, facing the Phe405 of the DFG motif, to an "out" conformation pointing away from the hydrophobic pocket. During the 1 μs trajectory, the reorientation of the αC-helix occurs with a period of 500 ns and a lifetime of 30 ns, while the flip of the Ile293 occurs approximately every 70 ns and has a lifetime of 200 ps. None of the two events have been observed in Src, where the αC-helix is much more stable and the orientation of the Leu317 side-chain remains unchanged for the duration of the whole simulation.

In summary, significant differences between Src and Abl sub-μs dynamics were found, with Abl showing a higher degree of flexibility. Moreover, the observed reorientation of both the αC-helix and residue Ile293 could favour the adoption of multiple conformations in Abl not accessible to Src.

3.2 PTmetaD calculation and free energy surface reconstruction

The observed changes in the dynamics of structural motifs surrounding the active site, prompted me to carry out a detailed and quantitative characterisation of the energetics of the DFG-flip, by means of well-tempered PTmetaD method. The choice of the collective variables (CVs) is of pivotal importance to assure a good sampling (see Materials and Methods, Section 2.1.3).

In order to sample the DFG-flip transition (DFG-in to DFG-out) four CVs able to discriminate between the two conformations were chosen (see Materials and Methods, Figure 2.4). In brief, three CVs involve residues of the DFG motif (CV1= distance <Asp404>-<Lys295>, CV2 = distance <F405>-<L317> for Src/I293 for Abl and

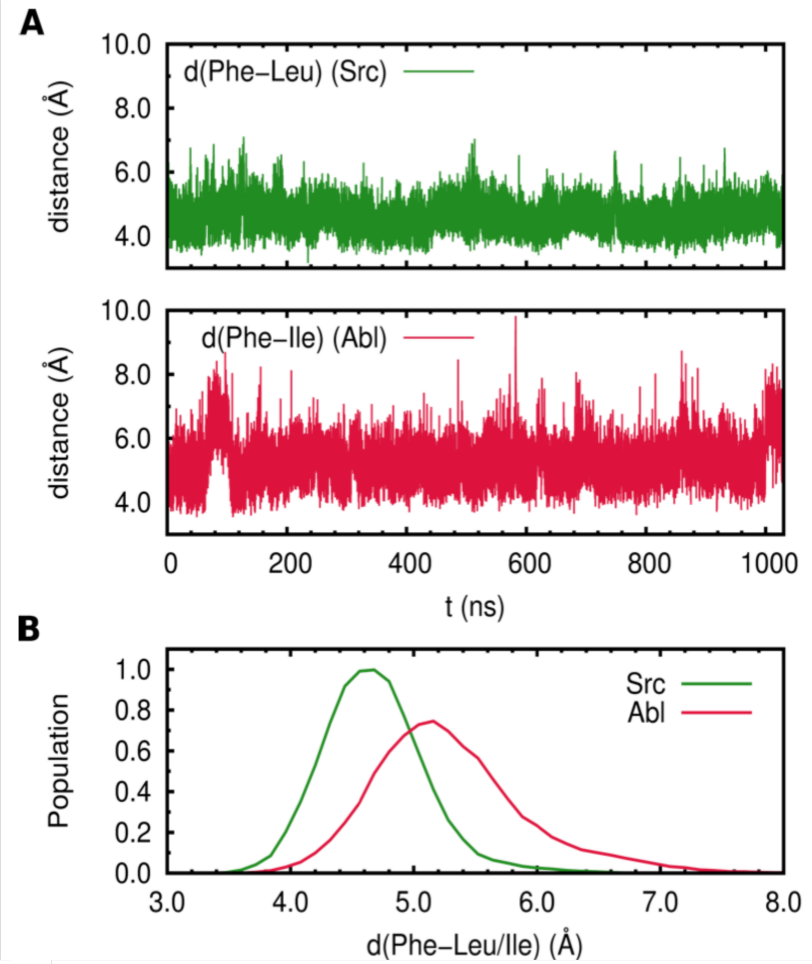

Figure 3.3: (A) Distances Phe-Leu of Src (in green) and Phe-Ile of Abl (in red) were monitored over the 1 μs long simulation. (B) Distribution of Src and Abl populations as function of the residue pair distance. Figure taken from Ref [187].

CV3 = a DFG dihedral combination), and the forth the N-terminal region of the A-loop (CV4 = distance <N381,...,H384>-<A408,...,I411>). The free energy surfaces (FESs) of Abl and Src have been reconstructed as a function of two CV1 and CV2, because those allowing a better sampling of the DFG flip transition.

Both simulations converged after 400 ns, as shown by the 1-D projection over time of CV1 and CV2 in Figure 3.4. Our choice of CVs resulted in FES where the DFG-in and DFG-out minima could be clearly distinguished (see Figure 3.5).

In both kinases the DFG-in conformation results to be the main minimum of the FES, identifying the more stable and populated conformation of basins 1 and 4 in Figure 3.5 for Src and Abl, respectively. The second more stable and pronounced minimum is the one corresponding to the DFG-out conformation (basins 3 and 8 in Figure 3.5). The free energy difference between the two conformations is 4.0 $\pm$ 0.5 kcal/mol for Abl and

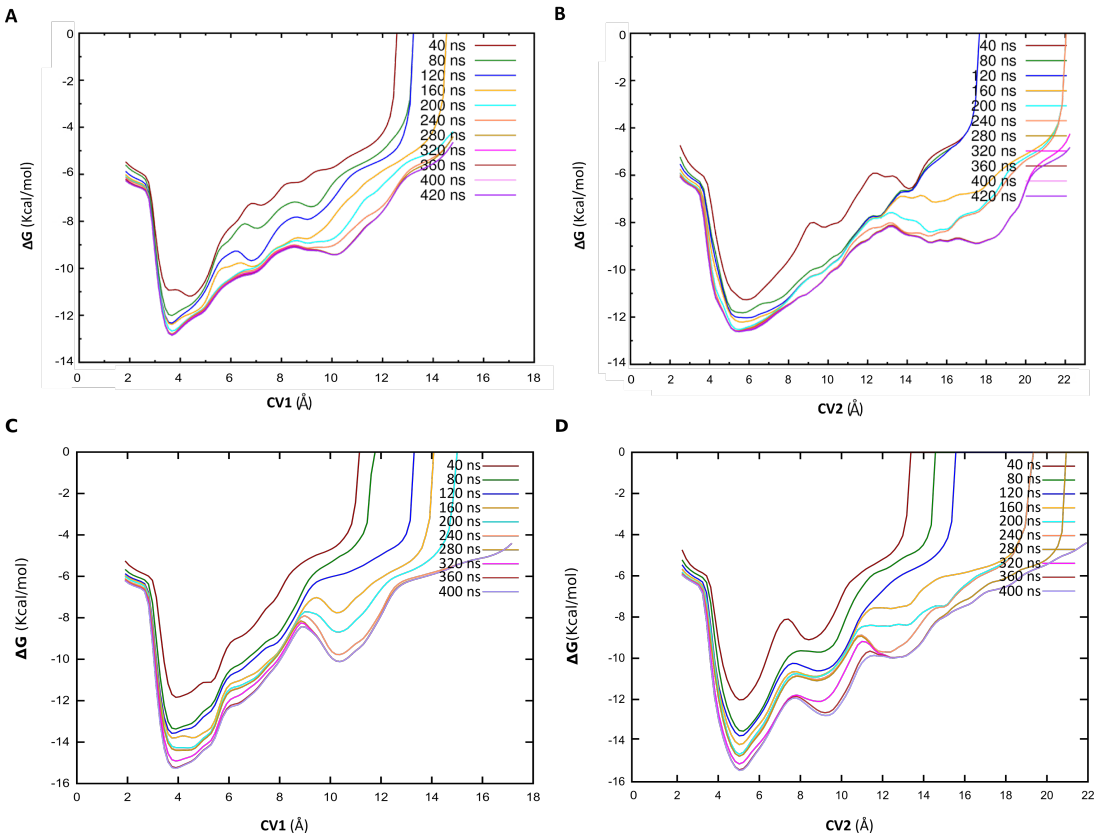

Figure 3.4: 1-D projection over time of the free-energy basins of CV1 (A) and CV2 (B) of Abl and Src (C, D). Figure taken from Ref [187].

6.0 $\pm$ 0.5 kcal/mol for Src. Thus, the DFG-out conformation in Abl is 2 kcal/mol more stable with respect to Src. Another important element differentiating the two FESs is the height of the energy barrier separating the two conformations, lower in Abl (6 $\pm$ 0.5 kcal/mol) than in Src (11 $\pm$ 0.5 kcal/mol).

The FES of Abl displays a third local minimum (basin 6 in Figure 3.5), just 1 kcal/mol higher than the DFG-out state. This minimum corresponds to an intermediate conformation in which the Asp404 is pointing downward (see structure 6 in Figure 3.5). This peculiar "Asp-down" conformation was first reported in the work of Shan *et al.*[70]. In total contrast, the corresponding area on the FES of Src exhibits a transition state instead of a minimum (state 2* in Figure 3.5). In the FES of Abl, the DFG-in and DFG-out conformations are connected via the "Asp-down" conformation and two transition states (5* and 7* in Figure 3.5). Thus, it is clear that the DFG flip mechanism taking place in Abl is significantly different from that of Src.

Furthermore, the local minima in the FES of Abl appear to be wider compared to Src, particularly the DFG-out minimum along CV2 (15-19 Å). This is a hint of a major

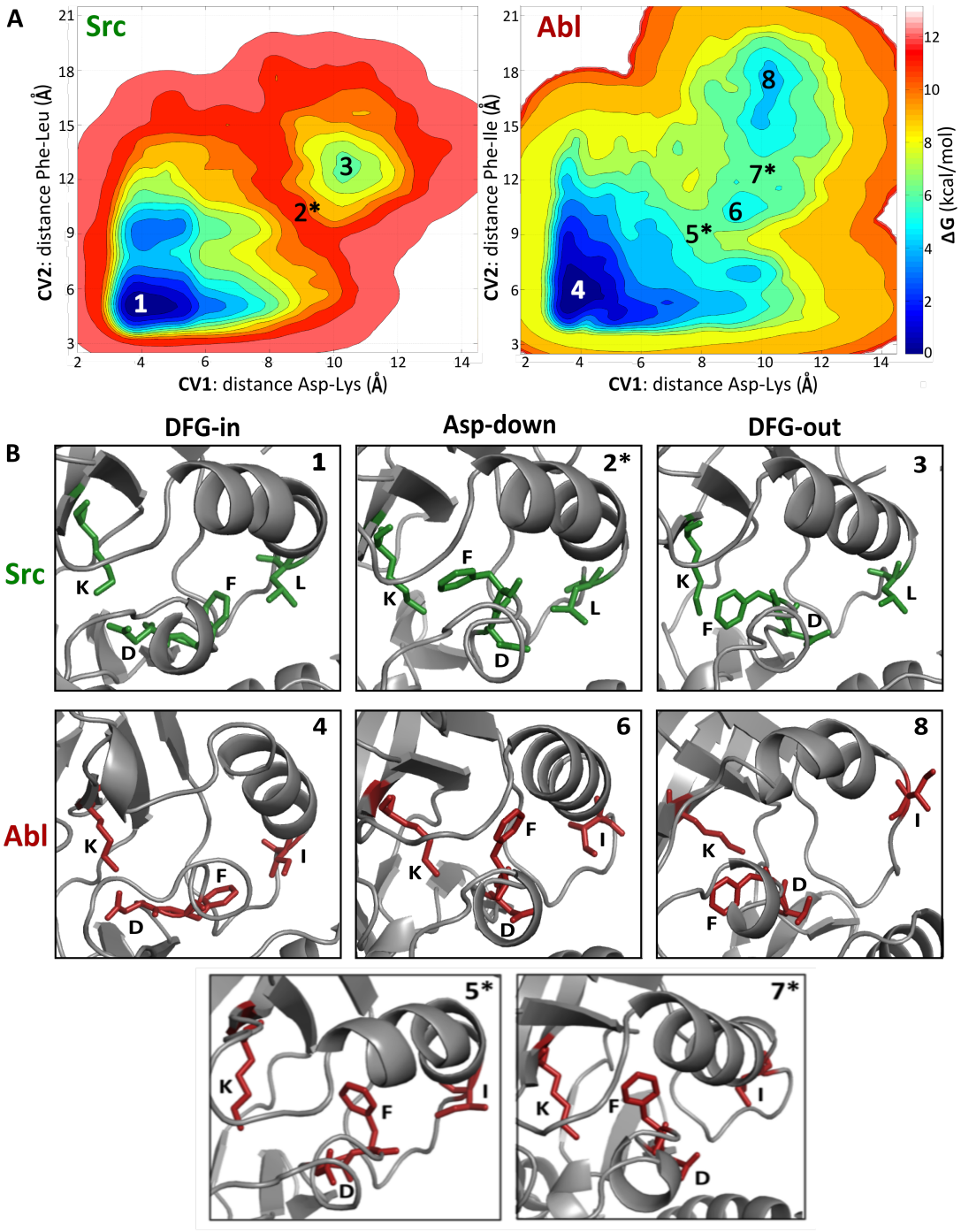

Figure 3.5: (A) Reconstructed FESs of Src (left) and Abl (right). The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. (B) Structures corresponding to the relevant minima and transition states. Figure taken from Ref [187].

structural plasticity along this variable. Thus, I extracted snapshots from the trajectory corresponding to the FE minima (for details see Section 2.1.8) and analysed their conformational features. The higher values reached by Abl along CV2, the Phe-Ile distance, are due to conformational changes in the αC-helix and in the orientation of residue Ile293 (structure 8 in Figure 3.5 and Figure 3.6). These results are in agreement with what previously observed, in the classical MD simulations, such as the higher values for the Phe-Ile distance and the Ile293 "flip". The enhanced flexibility of Abl has its distinctive mark along the Phe-Ile distance (CV2) while, the values of CV1 are comparable to Src. Thus, Abl has not only a more stable DFG-out druggable conformation,

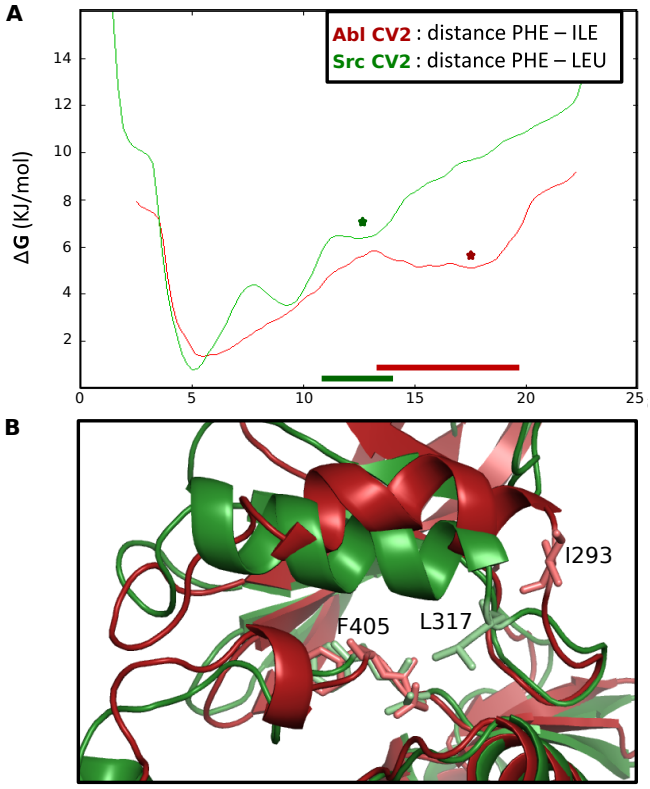

Figure 3.6: (A) Free energy profiles projected along CV2, distance Phe-Leu in green for Src and Phe-Ile in red for Abl. The green and red bars indicate the width of the respective minimum (B) Conformations adopted by the αC-helix of Src (in green) and Abl (in red), corresponding to the local energy minimum indicated with a green and red star, respectively, in panel A.

but also a lower barrier to overcome for the DFG flip to happen. Likely, the structural plasticity exhibited by Abl could have a major impact in lowering the height of the energy barrier.

3.3 Free energy reweighting

The higher flexibility and conformational changes identified in the N-lobe loops, αC-helix and A-loop region of Abl could result in a widening of the active site when adopting the DFG-out conformation (see Figure 3.7). A wider active site could favour the binding of imatinib to Abl, with respect to Src, facilitating drug engagement and minimising steric clashes.

To characterise in detail the conformational dynamics of the areas surrounding the active site, the reweighting technique was used to reprocess the free energy of Src and Abl, along four new CVs. The new CVs, not directly considered in the original PTmetaD simulations, described the movements of the N-lobe loops and αC-helix (see Materials and Methods, Section 2.1.8 and Figure 3.8, image A). The reweighted 1-D free energy profiles (see Figure 3.8) proved that, along all the new CVs, Abl explores a wider conformational space, with respect to Src, confirming its intrinsic structural plasticity.

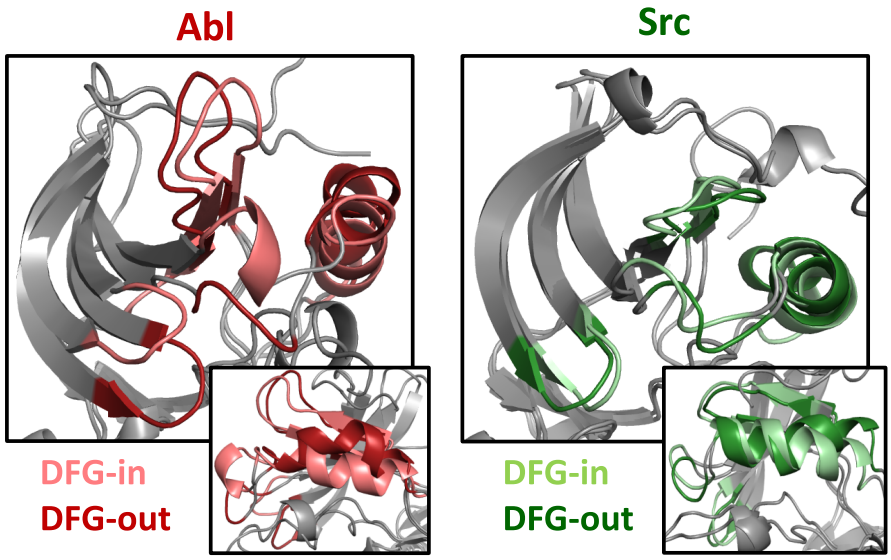

Figure 3.7: Alignment of structures extracted from the PTmetaD minima (DFG-in and DFG-out). Changes in the N-lobe loops and αC-helix conformations of Abl (in red) can be appreciated in the passage from DFG-in to DFG-out while, slight differences can be detected for Src (in green).

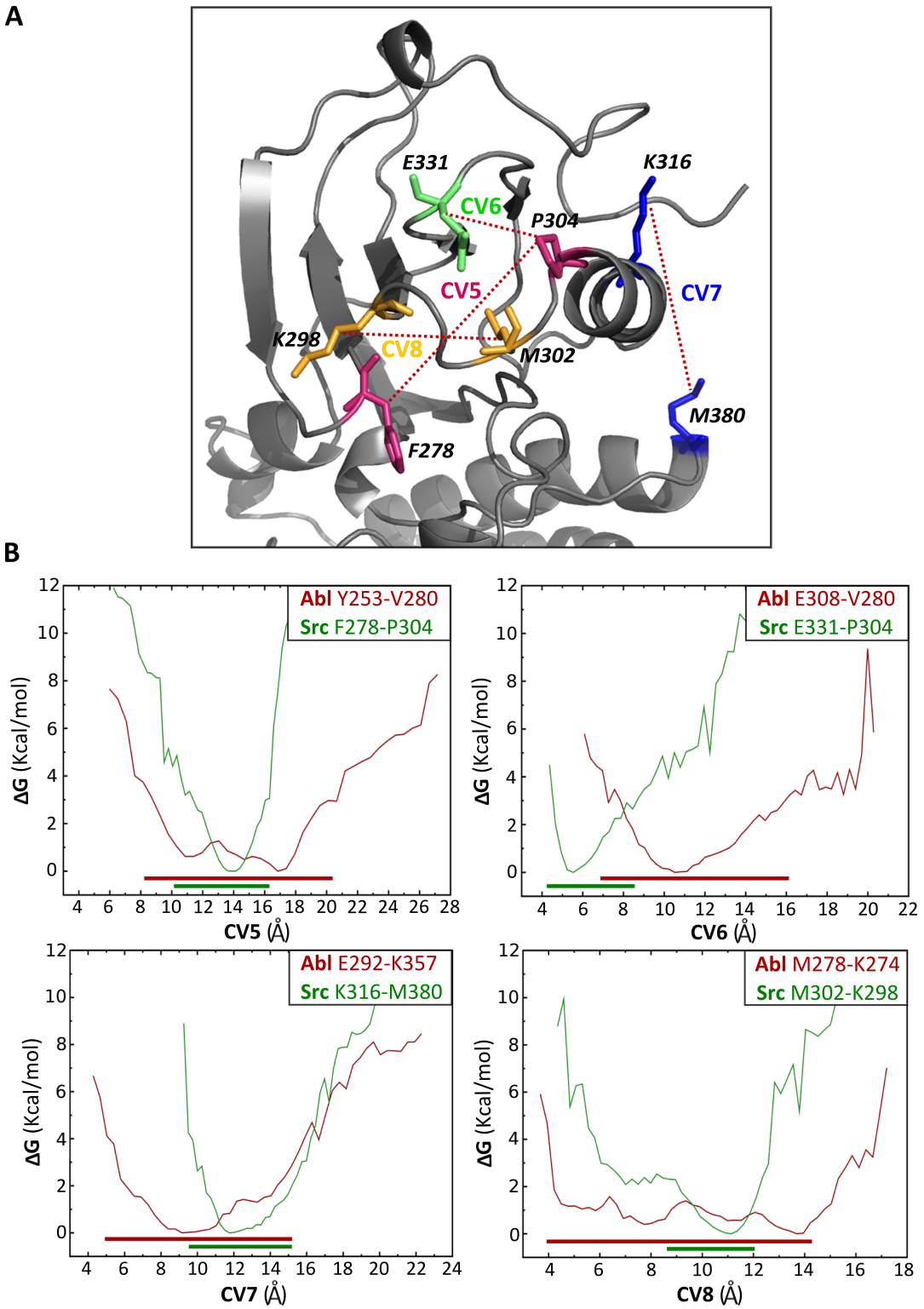

Figure 3.8: (A) Reweighting CVs (only Src numbering used). (B) 1-D projections of the reweighted FE profiles for Src (in green) and Abl (in red). The green and red bars underline the minima of Src and Abl, respectively. The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. Figure taken from Ref [187].

3.4 Solvent accessible surface area and water accessibility

The analyses of the classical MD and PTmetaD trajectories suggest that the active site of Abl opens up dynamically during the fluctuations of Abl in solution.

To address the hypothesis of a wider binding site in Abl than Src, the solvent accessible surface area (SAS) of the binding sites of both kinases was calculated (for details see Section 2.1.5). In the DFG-in conformation, the SAS is 5900 $\pm$ 200 Å^2 in Src and 6000 $\pm$ 100 Å^2 in Abl. In the DFG-out conformation, the active site in Src remains unaffected by the DFG flip (SAS = 5900 $\pm$ 100 Å^2) while in Abl the average size of the cavity increases to 6200 $\pm$ 200 Å^2 (see Figure 3.9).

The enlargement of the active site in the DFG-out conformation of Abl (Figure 3.9) should lead to an increase number of water molecules found in the vicinity of the DFG motif. To perform the count of water molecules within 5 Å of the DFG residues,

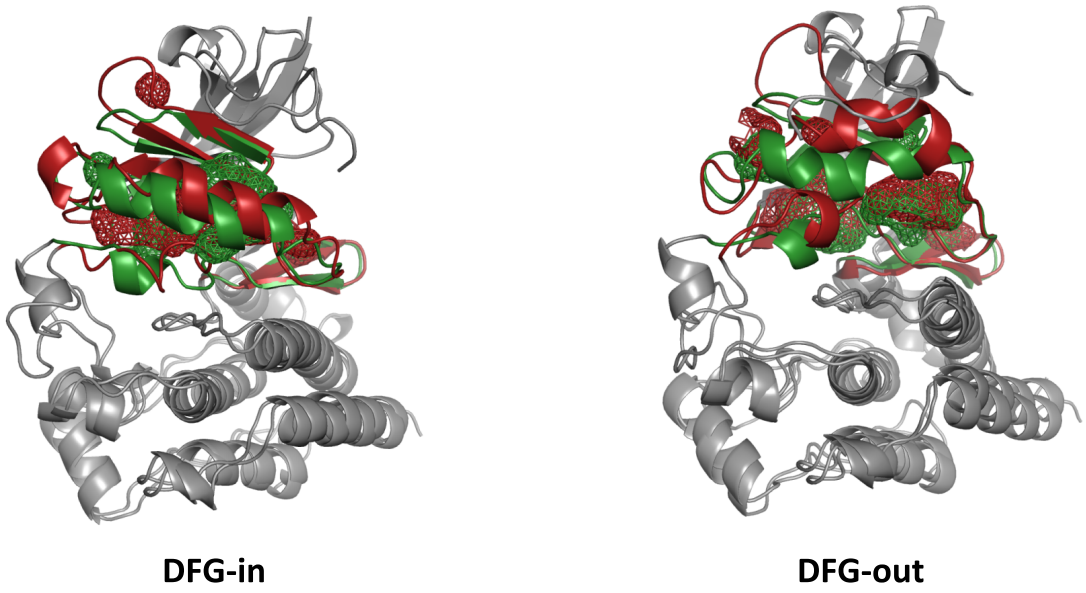

Figure 3.9: Comparison of the cavity dimensions in the DFG-in and DFG-out conformations of Src (in green) and Abl (in red).

new MD simulations of 20 ns each for the DFG-in and DFG-out of Src and Abl were performed (see Section 2.1.5). In Src the number of water molecules remains unaffected in both conformations (19 $\pm$ 3). In comparison, the averaged number of water molecules in Abl is increased due to the entrance in the cavity of up to three addi-

tional water molecules in the DFG-out conformation (19 ± 5 in the DFG-in; 22 ± 5 in the DFG-out). In Abl, the additional water molecules surrounding the Asp404 residue could contribute to the stabilisation of the inactive conformation, thanks to a more favourable environment and the formation of hydrogen bonds (see Figure 3.10).

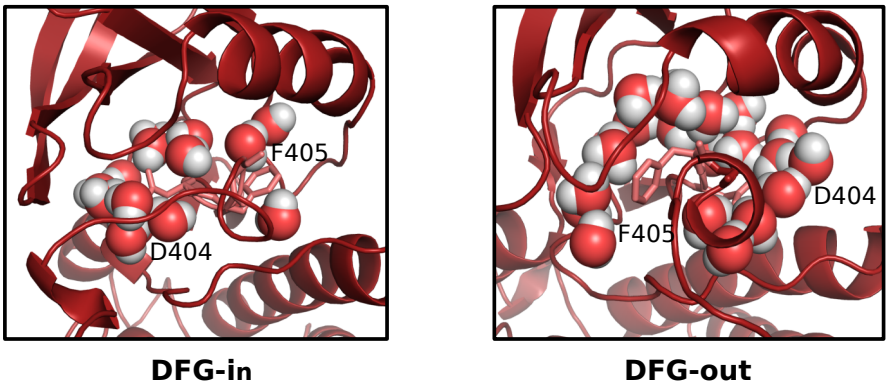

Figure 3.10: Water molecules within 5 Å of the DFG motif of Abl (shown in sticks) in the DFG-in and DFG-out conformation.

3.5 Calculation of the pK$_a$ and influence of Asp404 protonation

As discussed in the Introduction (see Section 1.4.1), the binding of imatinib to Abl was proposed to be pH dependent. This might be due to variation in the pK$_a$ of residue Asp404 in the DFG-in and DFG-out conformation[70]. To quantify the contribution of pH to the flip, the pK$_a$ of the Asp404 residue was calculated for the DFG-in and out conformations in both kinases. The average pK$_a$ values for Abl were 3.4 and 4.4 for the DFG-in and DFG-out, respectively, meaning a positive contribution to the ΔG of the transition ($\Delta\Delta G_{in-out}$ = 1.4 ± 0.9 kcal/mol). In the case of Src, the average pK$_a$ were 4 in the DFG-in and 2.6 in the DFG-out. Therefore, the adoption of the DFG-out conformation is destabilised by almost 2.0 ± 0.9 kcal/mol.

3.6 Discussion

The DFG-out imatinib-binding conformation is not only energetically more stable in Abl than in Src, but the kinetics of the DFG transition is also more favourable due to a lower free-energy barrier to cross. This is in agreement with the IC$_{50}$ experiments that proved imatinib to be a strong inhibitor of Abl but not of Src[65]. The better

accessibility of the DFG-out conformation could be associated with a higher plasticity of the Abl structure, as seen by the analysis of the sub-μs dynamics.

Furthermore, in Abl, Asp404 is more solvent exposed in the DFG-out conformation, compared to Src, due to reorientation of the αC-helix and enlargement of the binding site. This conformation facilitates the entry of additional water molecules into the cavity and thus stabilises Asp404. Moreover, it may also facilitate the entry of imatinib into the cavity.

The pK_a analysis shows a pH dependence of the binding of imatinb as well as of the DFG-out population. At physiological pH (cytosol 7.2), according to the calculated pK_a values, the Asp404 of both kinases would be in the deprotonated form, thus drastically reducing the population of the DFG-out in Abl (pK_a DFG-out = 4.4) and even more in Src (K_a DFG-out = 2.6).

Overall, the differences in the kinetics and thermodynamics of the DFG transition can explain the different imatinib activity towards Src and Abl, also confirming a conformational selection binding mechanism of imatinib[187, 210, 211].

Chapter 4

The correlation of sub-μs dynamics to drug binding in tyrosine kinases

In the previous chapter, the sub-μs dynamics was seen to influence the conformational dynamics related to the DFG flip transition of Abl and Src. The greater flexibility shown by the Abl kinase domain (KD) seems not only to facilitate the adoption of the druggable conformation, but also to increase the size of the binding site in the DFG-out. Together these effects could positively affect the binding of the drug imatinib and explain its higher affinity for Abl.

To further confirm the relation existing between the sub-μs dynamics of the tyrosine KD and the imatinib activity, in this chapter I have extended the analysis of the sub-μs dynamics to a larger pool of tyrosine kinases. As discussed in the Introduction (Section 1.4.1), imatinib is not so selective as it was initially thought[61]. It effectively inhibits several TKs, making it useful in the treatment of several cancers beyond chronic myeloid leukemia (such as gastrointestinal stromal tumour (GST) and metastatic tumours)[212]. However, imatinib also bind to other TKs but showing very reduced activity[212].

Classical molecular dynamics simulations (MD) of several TKs were performed. The TKs included in the study were chosen on the basis of their sequence homology to Src and Abl, and of their different sensitivity towards imatinib (measured by IC_{50}).

4.1 Selection of the tyrosine kinases included in the study

The selected TKs share at least 45% sequence identity to both Src and Abl, and they are inhibited by imatinib with varying strengths (refer to Table 1.1 of the Introduction). They are: Kit (IC_{50}=0.31 μM[63]), Lck (IC_{50}=9 μM[56]), Met and EGFR (IC_{50}>100 μM[65, 66]). The IC_{50} is a quantitative measure of how effective is a drug. Higher is the IC_{50}, lower is the inhibitory effect, activity, of the drug. It is a measure of the drug concentration needed to inhibit 50% of the substrate available.

Kit is the most sensitive kinase to imatinib after Abl, while Lck shows a discrete sensitivity despite the huge sequence identity to Src (up to 70%). On the contrary, both Met and EGFR are not effectively inhibited by imatinib even if, as in the case of Src, they can adopt inactive DFG-out conformations and imatinib shows the same binding mode as seen for Abl[64, 65, 66, 60, 64, 63, 56].

4.2 Allosteric coupling analysis

As discussed in Section 1.2.1, tyrosine kinases are characterised by a flexible structure, which is subject to significant conformational changes involving several conserved structural elements. The concerted motions of these elements, leading to the adoption of different conformations, require some degree of correlation.

To start understanding the allosteric connection of remote sites of the KD to the binding site, I used a normal modes analysis approach (NMA) combined with a first-order perturbation theory[178, 180] (see Section 2.1.5). The resulting allosteric coupling maps are shown in Figure 4.1.

A high cross-correlation of the N-lobe β elements and loops (res. 1-45 in Figure 4.1) is observed in all kinases, as a consequence of the stable β-sheet secondary structure. Only for a subset of them, namely Lck, EGFR and to a lesser extent Kit and Met, this network is extended to include the αC-helix and the preceding β3-αC loop (res. 45- 60 in Figure 4.1). These kinases (Lck, EGFR, Kit and Met) are also the ones exhibiting a much higher correlation between the smaller β4-β5 hairpin (res. 70-80 in Figure 4.1) and the rest of the N-lobe. The central region (res. 80-150 in Figure 4.1), encompassing the linker between the N-lobe and the C-lobe and part of

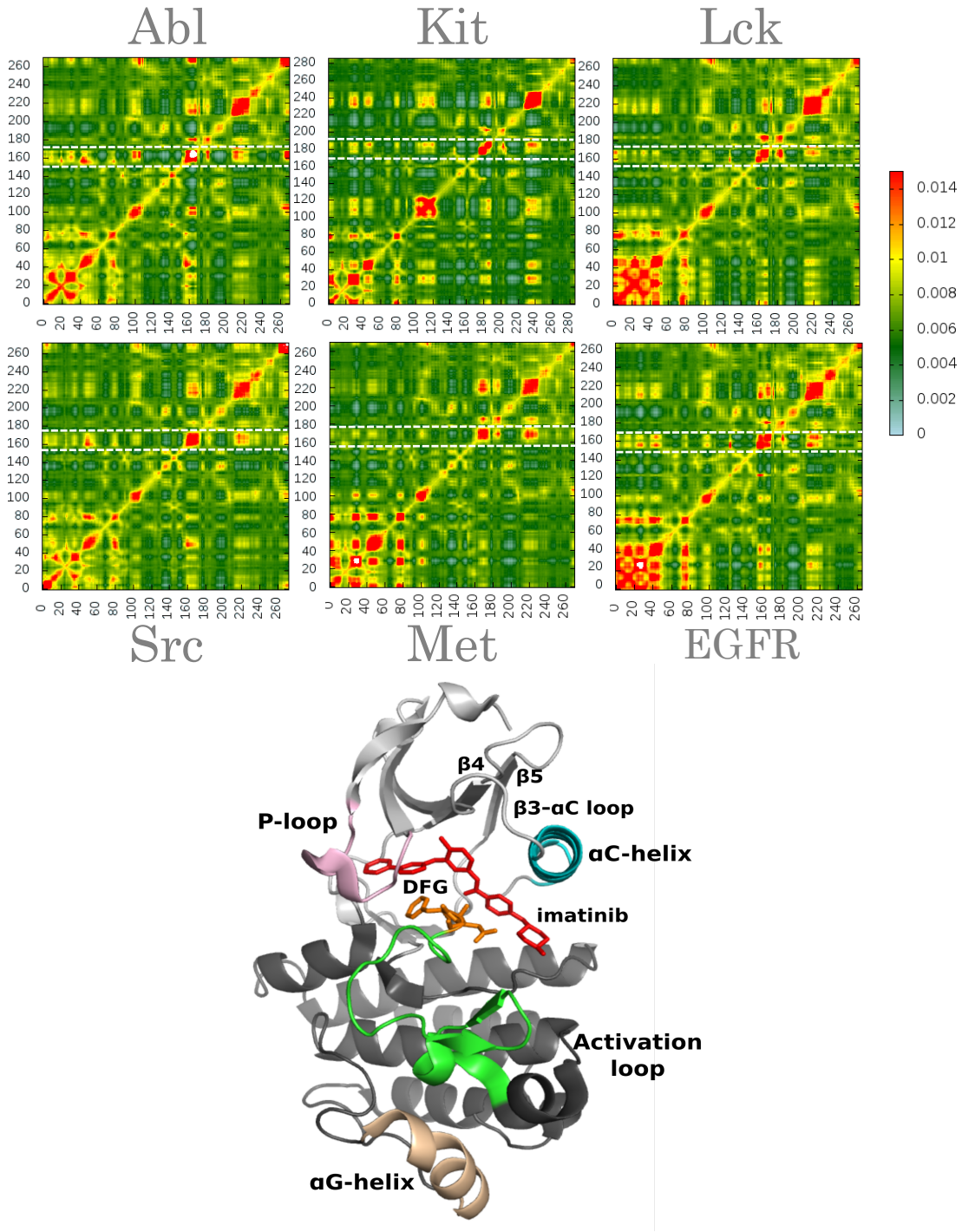

Figure 4.1: Allosteric coupling maps of Src, Abl and the homologous TKs together with the structure of the kinase domain with highlighted the elements involved in the allosteric coupling. In the maps, the A-loop region corresponds to the area included in the white dashed lines. Figure taken from Ref [93].

the hydrophobic ATP binding pocket, shows negligible couplings with other structural elements.

Interestingly, inter-lobe couplings are observed for the activation loop (res. 150-170 in Figure 4.1). The A-loop plays a key role in the conformational changes occurring in TKs[73, 17], and is involved in the transition from the active (A-loop open) to the inactive conformation, in which the loop is predominantly closed. Abl shows a strong correlation between the A-loop, the αC-helix, the P-loop and all the N-lobe loops in general. This network is less prominent, but still present, in the other kinases analysed, as Src, Kit and EGFR. In the case of Src, the A-loop movement results mainly coupled with the αC-helix and the preceding loop. Conversely, Lck and Met show particularly poor couplings for the A-loop (see Figure 4.1). The main element of the C-lobe involved in allosteric coupling is the αG-helix, showing a correlation with the A-loop movement in all kinases. This correlation is particularly prominent in Kit and EGFR, where it extends partially to the N-lobe as well (see Figure 4.1).

Taken together, the allosteric coupling analysis predicts a highly correlated N-lobe, where many structural elements move in unison, and a less correlated C-lobe, where only the A-loop and the αG-helix emerge as correlated. What is more, the few non-trivial inter-lobe correlations all involve the A-loop, whose motion appear as the cornerstone of the large-scale allosteric connections in the kinases network, leading inter-lobe movements. The analysis not only confirms TKs as proteins characterised by a complex allosteric network[17, 15], but also identifies Abl as the kinase with the greater correlation among all the structural elements important for the protein conformational changes.

4.3 Classical MD simulations and RMSF analysis of the homologous TKs

Long atomistic MD simulations were performed for all the kinase domains (KD) of the chosen TKs, adopting the DFG-in conformation (see Section 2.1.4). To investigate the sub-μs dynamics of the homologous TKs, with respect to Src and Abl, the root mean square fluctuation analysis (RMSF) was performed (see Section 2.1.5). As it can be seen in Figure 4.2 (bottom panels), the N-lobe loops and the A-loop are confirmed as

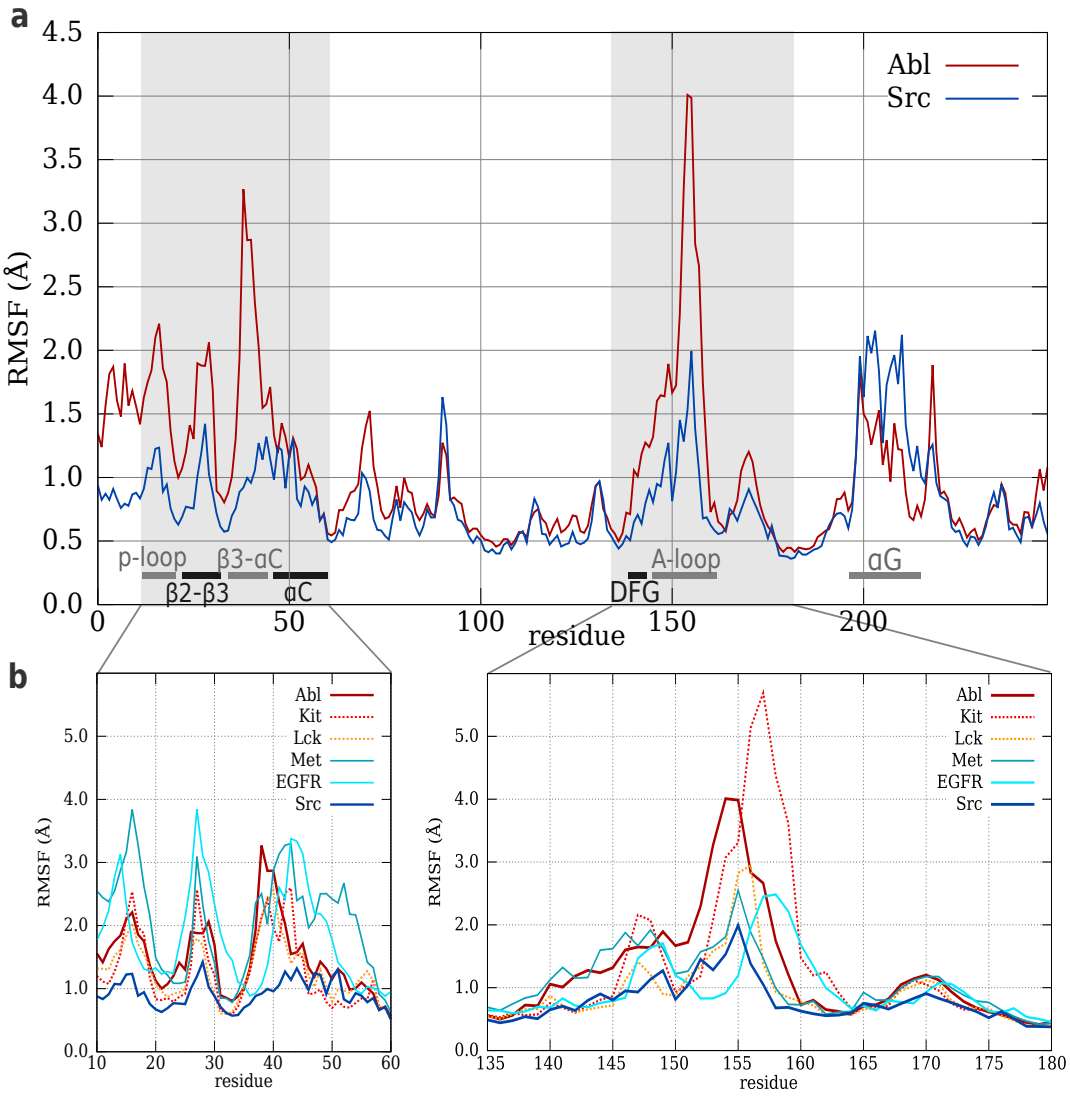

Figure 4.2: (a) RMSF profiles of Src (blue) and Abl (red). Grey panels identify the N-lobe and the A-loop region. (b) RMSF profiles of the homologous TKs. In the two panels are reported the RMSF profiles of the N-lobe region (left) and A-loop region (right), being the areas whose flexibility varies most among the pool of TKs. Shades of red identify the flexible TKs, shades of blue the more rigid ones. Dotted lines are used for clarity. Figure taken from Ref [93].

the most variable regions of the structure, also in the homologous TKs.

For what concerns the A-loop, Lck and Kit reach or even exceed the RMSF values registered for Abl (peaks at 4.2 Å, 3.0 Å and 5.5 Å for Abl, Lck and Kit, respectively, see Figure 4.2), while Met and EGFR have a more rigid profile in this region, more similar to the one observed for Src (RMSF values of 2.0 Å, 2.6 Å and 2.5 Å for Src, Met and EGFR, respectively, see Figure 4.2). Looking at the N-lobe region, again Src has the most rigid profile, due to an RMSF value of the N-lobe that never exceeds 1.5 Å. Higher values are observed for Abl, Lck and Kit, in a range that goes from 1.8 Å to 3.2 Å (see Figure 4.2). Surprisingly, Met and EGFR report the highest RMSF values in the N-lobe, above 2.5 Å (Figure 4.2). In the αG-helix region (see Figure 4.3), EGFR is the only TK showing a flexibility comparable to the one of Src (2.0 Å).

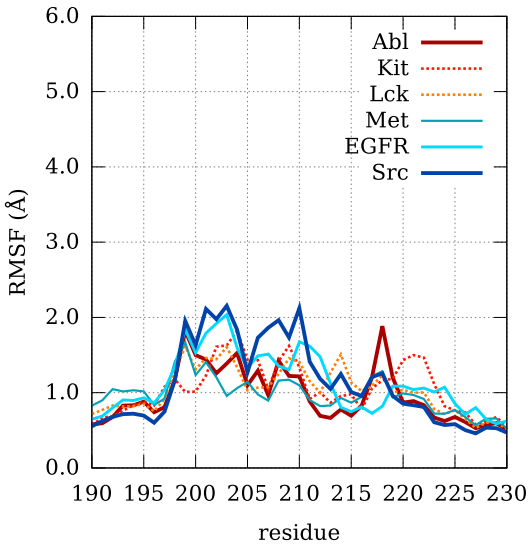

Figure 4.3: RMSF profile of the αG-helix of Src, Abl and the homologous TKs. Shades of red identify the flexible TKs, shades of blue the more rigid ones. Dotted lines are used for clarity. Figure taken from Ref [93].

While the flexibility of the N-lobe seems to vary across all kinases without a particular apparent trend, there is a remarkable (anti)correlation between the A-loop flexibility and the imatinib IC_{50} (see Table 4.1). The kinases having the most flexible A-loop (Abl, Kit and Lck) are also the ones with the lowest IC_{50} value, meaning that imatinib is able to efficiently inhibit the kinase. On the contrary, the kinases with a lower RMSF in the A-loop show a higher IC_{50} in the literature.

The N-lobe and the A-loop are known to be involved in the conformational changes leading to the adoption of the druggable, DFG-out, conformation[20]. The

higher flexibility of both these regions, as in Kit and Lck, could facilitate the kinase conformational transition (DFG-in to DFG-out), as previously proved for Abl (Chapter 3). However, when this condition is not satisfied, as for Met and EGFR (flexible N-lobe only), the imatinib activity results compromised. From this analysis, was confirmed the relation existing between sub-μs dynamics, conformational dynamics and imatinib activity in TKs.

TK	A-loop RMSF (Å)	IC_{50} (μM)
Abl	4.2	0.037 [62]
Kit	5.5	0.3 [63]
Lck	3.2	9 [56]
Met	2.6	>100 [64, 65]
EGFR	2.5	>100 [66]
Src	2	>100 [66, 65]

Table 4.1: RMSF of the A-loop region and IC_{50} values of imatinib reported in literature for Abl, Src and the homologous TKs[93].

4.4 Principal component analysis

All tyrosine kinases are characterised by dominant motions involving the N-lobe and the C-lobe of the kinase domain. Two main motions have been described[70, 41]: the hinge-bending motion, where N-lobe and C-lobe get closer (see Figure 4.4, image a), and the lobe twist of the N-lobe relative to the C-lobe (see Figure 4.4, image b).

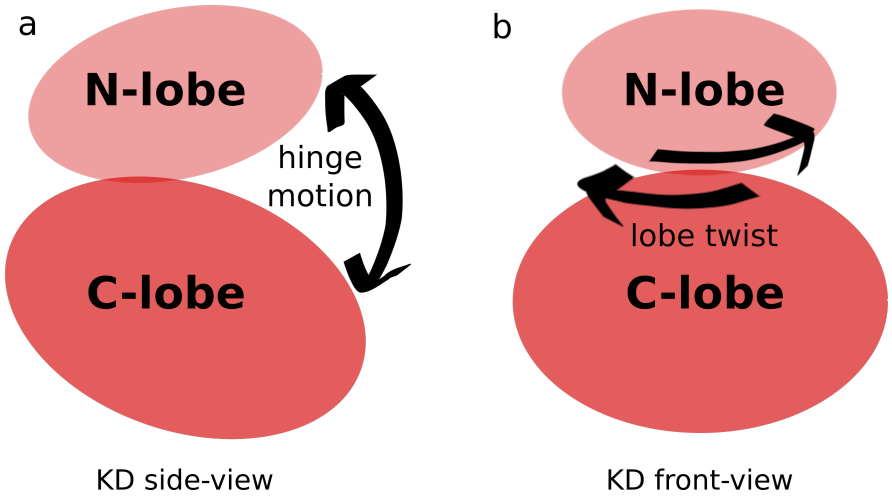

Figure 4.4: Schematic representation of the KD dominant motions: the hinge-bending motion (a) and the lobe twist (b). Both are characterised by movements of the N-lobe relative to the C-lobe.

To investigate if the change in the sub-μs dynamics could also affect the kinase dominant motions, a PCA analysis of the TKs trajectories was performed. The first two eigenvectors contributing most significantly to the enzyme dynamics were chosen. As discussed in Section 2.1.5, the trajectories of all homologous TKs were projected on Abl first and second eigenvectors. This allowed to compare the conformational space visited by every kinase with respect to Abl (chosen as reference of a flexible kinase in the study). The first eigenvector describes the movement of the N-lobe (hinge motion), while the second eigenvector describes mainly the lobe twist (see Figure 4.4).

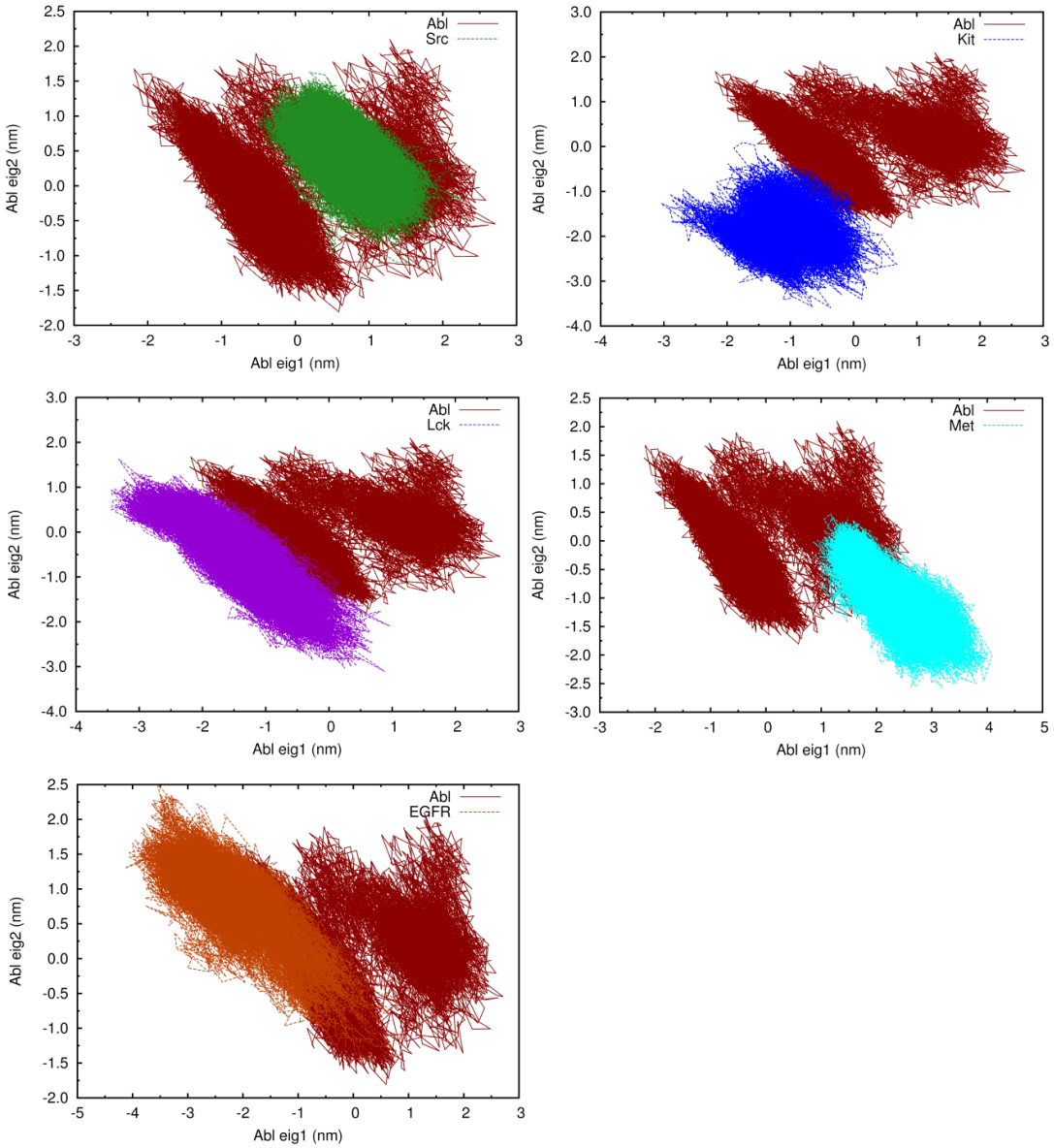

Figure 4.5: PCA analysis. Comparison of the conformational space explored by the homologous TKs under study (Src, Kit, Lck, Met and EGFR) with respect to Abl. The first eigenvector (Abl eig 1) describes the hinge motion, while the second eigenvector (Abl eig 2) the lobe twist.

All the TKs under investigation explore a smaller subspace compared to Abl (see Figure 4.5). Moreover, most TKs (Kit, Lck and Met) visit a different portion of the conformational space, in particular considering the first eigenvector, which corresponds to the hinge motion (see Figure 4.5). The projections of Src and EGFR, instead, partially overlap with the one of Abl (see Figure 4.5).

As expected, TKs share a common dynamics due to their KD structure, but they have in addition a unique dynamic pattern determined by their own sequence. The plasticity shown by Abl is particularly high, and it spans a great number of different conformations. For example, the N-lobe and the αC-helix of Abl have a peculiar behaviour, previously described in Chapter 3, not observed in the homologous TKs thus, becoming a distinctive feature of Abl.

4.5 Experimental validation

On the whole, the results on the homologous TKs of Src and Abl show a link between the A-loop flexibility and the activity of imatinib: the kinases with the most flexible A-loop are the ones reporting the lower IC_{50} values. However, other structural elements, such as the N-lobe loops, might also play a role in the drug binding.

In the Introduction (see Section 1.5.2) I discussed the differences in the binding modes of two classes of kinase inhibitors: type I and type II[115, 108]. Type I inhibitors occupy just the ATP binding site, docking below the P-loop (refer to Figure 1.20 of the Introduction). Type II, as imatinib, occupy both the ATP binding site and the allosteric pocket (refer to Figure 1.20). In the light of this, the high flexibility of the P-loop region of Met and EGFR shown in the RMSF analysis gains importance[115, 108].

In this section, the experimental results in Ref [213] were used to cross-validate the model of kinase dynamics proposed in this thesis, against a broader range of TK inhibitors. The aim was to better understand the correlation existing between the sub-μs dynamics of the N-lobe and A-loop regions and the binding of drugs.

Peterson and coworkers[213] carried out a remarkable wide-scale study on the catalytic activity of various kinases against commercially available inhibitors. The results of this analysis are reported in Figure 4.6. Remarkably, the 'activity matrix' across the six selected kinases for the subset of compounds clustered itself in three different blocks, covering around the 70% of the analysed data. The first block is characterised

by a high activity towards four kinases (Abl, Kit, Lck and Src) and comprises type I inhibitors, as dasatinib and sunitinib. The second block differs from the first one for the lack of activity towards Src, while retaining the effect against Abl, Kit and Lck. This block comprises imatinib itself and other compounds as nilotinib and masitinib (type II inhibitors). A third block is observed, with a strikingly different pattern, composed mainly of drugs active against EGFR only. Not surprisingly, the well known highly specific EGFR inhibitors erlotinib and gefitinib were found in this category. The remaining compounds, not belonging to any group in particular, were identified as either highly selective towards one kinase (e.g. Src and Met inhibitors) or as nonspecific inhibitors, like staurosporine.

When compared to the flexibility pattern of the A-loop and of the N-lobe, significant analogies could be observed. The first block of compounds presents a separation across the pool of kinases similar to the one observed for the N-lobe RMSF. Abl, Kit, Lck and Src with lower N-lobe fluctuations have also a lower residual catalytic activity, so that are effectively inhibited when treated with compounds of this group (type I). Conversely, the second group (type II) has a trend which is opposite to the flexibility observed for the A-loop. Abl, Kit and Lck with a more flexible A-loop have a lower residual activity.

Therefore, it seems that a low P-loop flexibility is needed, probably to assure a better stability to the binding of type I inhibitors. As evidence of this, type I inhibitors are not effective in kinases showing an excessive P-loop flexibility, as Met and EGFR. On the contrary, the high flexibility of the A-loop region seems to be crucial for the activity of type II inhibitors, since all kinases having a rigid A-loop are not inhibited by imatinib or any other type II binder.

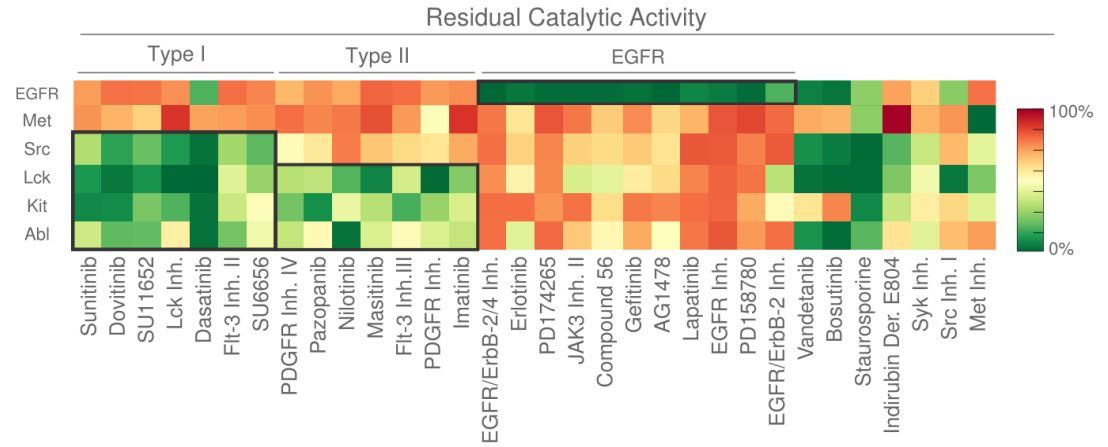

Figure 4.6: The activity matrix shows the residual catalytic activity for each kinase-drug match retrieved from the results obtained in the work of Ref. [213]. Shades of green identify the kinases effectively inhibited by the considered drug, so having a low residual catalytic activity, while shades of red identify kinases not inhibited by the drug and maintaining an high residual activity. With residual activity is meant the ability of the kinase to phosphorylate biological substrates. The drugs tested in the study have been grouped according to their binding mode to the kinase (type I, type II). To the other groups belong the EGFR specific inhibitors and the nonspecific inhibitors).

4.6 Discussion

Tyrosine kinases, even those with a significant sequence identity, show a wide range of sensitivity towards inhibitors that cannot be easily explained considering only their sequence or the 3D structure. The results presented in this Chapter show that sub-μs dynamics is able to discriminate among kinases, even when their sequence is more than 45% identical, thus becoming crucial as a descriptor of the kinase behaviour, as a kinase 'fingerprint'. Every TK turns out to have a peculiar sub-μs dynamics, that could be related to the inhibitor's selectivity.

The engagement and final binding of a drug to a target is a complex process, that involves many steps and conformational changes in both the target and the drug. Knowing the structural and also the dynamical specificities of a target is therefore essential for the design of effective drugs. Nowadays, the view of the binding event has evolved from the "lock and key" theory[4]. The targets are considered as dynamic entities and more emphasis is given to the study of their intrinsic dynamics and conformational landscape[3, 9, 10].

In this Chapter a clear trend was identified, proving that the dynamics of specific regions in the TK structure are more involved than others in the binding process of

drugs. For example, the dynamics of the P-loop and the A-loop is strictly correlated with the activity of inhibitors, because both involved in the conformational changes that lead to the adoption of the druggable conformations (DFG-in for type I inhibitors, DFG-out for type II).

Chapter 5

Src mutagenesis confirms the link between flexibility and imatinib binding affinity

Based on the computational evidence discussed in Chapters 3 and 4, the flexibility of specific regions within the kinase domain seems to be connected to imatinib activity in TKs. To further validate these results, a number of mutants were engineered in order to increase the sub-μs dynamics of the Src KD.

In this Chapter, I will describe the mutagenesis study undergone to understand whether inducing a change in the kinase dynamics is sufficient *per-se* to increase the sensitivity towards imatinib. In the first part of the mutagenesis study, MD simulations of the engineered mutants of Src were performed in order to predict their flexibility. Subsequently, those showing a significant increase in flexibility *in silico* were chosen for the second, experimental, part of the study. The selected mutants were expressed in *E.coli* cells, then purified and finally the binding affinity (K_d) of imatinib was determined by means of isothermal titration calorimetry (ITC).

The identification of mutants able to alter the sub-μs dynamics of Src, as well as their expression and purification, required a lengthy trial and error procedure.

5.1 Selection of the Src mutants included in the study

The amino-acid sequence of Abl and Lck, both flexible and imatinib sensitive kinases, were taken as references. Since Lck shares the highest sequence identity with Src (70 %), it was used as template, reducing the number of residues to mutate.

The mutations were designed in order to increase the structural flexibility of Src, following two main strategies. On the one hand, mutating regions with a major variation in dynamics among the homologous TKs, on the other hand, breaking contacts responsible for structural stabilisation.

The mutations aimed to modify Src flexibility were engineered in the P-loop, β3-αC loop, A-loop and αG-helix (see Figure 5.1). In these regions, the residues of Src, that differ among the three kinases, were mutated into the corresponding residues of Lck and Abl (to obtain a progressive increase in flexibility). The selected residues were: Q275 (P-loop), P299 (β3-αC loop), Q420 (A-loop) and V461 (αG-helix). To note is that the mutation of the αG-helix was meant to rigidify this region (and not to make it more flexible), in order to mimic Abl behaviour.

Since this approach was not fully successful, other mutants were designed (see Figure 5.1), this time aimed to break crucial contacts within the Src structure: E415G (A-loop), Y416G (A-loop) and M302G (αC-helix). Single and triple mutants of Src were designed.

5.2 MD simulations and RMSF analysis of the engineered mutants of Src

To test the flexibility of the Src mutants and determine the impact that mutations have on the sub-μs dynamics, preliminary MD simulations were performed (for details see Section 2.1.4). Four single mutants were initially tested (see Figure 5.1): Q275A, P299Q, Q420E and V461I (Src to Lck mutations).

In general, every engineered mutant perturbs the dynamics of regions far away from the mutated site. This confirms the presence of an extensive allosteric network of coupled residues spanning the kinase domain, as also mentioned in Ref [17].

As shown in Figure 5.2, the regions most affected by the mutations are: the P-loop, the αG-helix and the αD-helix. The mutations (see Figure 5.1) increase the flexibility

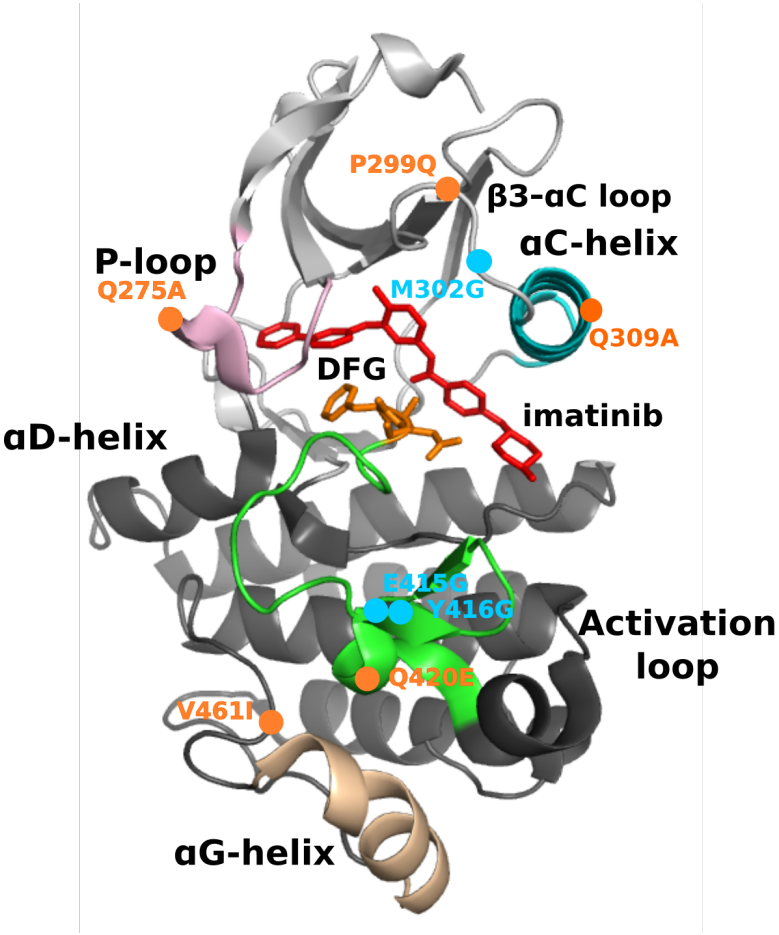

Figure 5.1: Structure of the kinase KD with highlighted the positions of the engineered mutants of Src. The mutations affecting Src dynamics are coloured in orange, the ones affecting Src contacts are coloured in cyan.

of the P-loop and decrease the ones of the αG-helix. The only exception is P299Q (β3-αC loop mutant), in which the αG-helix becomes more flexible than in the wild-type (3.8 Å, Figure 5.2).

The mutagenesis analysis shed the light on the strong allosteric correlation between the P-loop and the αG-helix, not previously detected (allosteric coupling analysis, Section 4.2). A clear example of this connection is the αG-helix mutant (V461I), in which the P-loop is the only region affected by the mutation (see Figure 5.2). Finally, the behaviour shown by the αD-helix is unusual, and not further observed in the RMSF analyses performed on the triple mutants of Src engineered later on.

As just mentioned, since the increase in the sub-μs dynamics of Src WT obtained by single mutations was minimal, triple mutants were attempted. The triple mutants tested were Q275A/P299Q/Q420E (mutations Src to Lck) and Q275G/P299E/Q420A

(Src to Abl). The mutated areas were the P-loop, the β3-αC loop and the A-loop (see Figure 5.1).

Taking the RMSF profiles of Src, Lck and Abl as references for a rigid, intermediate and highly flexible KD, the triple mutants are predicted, by the simulations, to mostly affect the flexibility of the N-lobe (see Figure 5.3 and Figure 5.4). Q275A/P299Q/Q420E reaches the 2 Å flexibility in the P-loop and in the β2-β3 loop. Unfortunately, the flexibility of the A-loop region seems to be unaffected by the selected mutations (see Figure 5.3), while the αG-helix shows a decreased flexibility of about 1 Å, as in Abl (see Figure 5.4). A similar flexibility pattern is predicted for the Q275G/P299E/Q420A mutant. As expected, the average intensity of the fluctuations is higher and in some regions, as the P-loop and the αC-helix, exceeds the values predicted for Abl (see Figure 5.3). A slight increase in flexibility results for the A-loop. However, even a triple mutation seems not to be sufficient to affect dramatically the dynamics of the A-loop. This might be due to the allosteric regulation acting on this functionally relevant motif.

To test this hypothesis, I also tried to affect the flexibility of the A-loop indirectly by mutating the αC and αG helices, supposed to be allosterically connected to the A-loop[17]. The new triple mutants (see Figure 5.1) created were Q275A/P299Q/Q309A (Q309A, mutation on the αC-helix Src to Lck) and Q275G/P299E/V461S (V461S, mutation on the αG-helix Src to Abl). Unfortunately, no appreciable change in the flexibility was noticed for the first mutant, probably due to the very high sequence variability of the αC-helix region throughout kinases, making difficult obtaining an effect only with a single point mutation. Better results were obtained with the αG-helix mutant (see Figure 5.3), where the P-loop and the αG-helix itself were flexibilised and rigidified respectively, resembling the Abl behaviour (average RMSF of 1 Å in the αG-helix, see Figure 5.4). However, again no effect on the A-loop flexibility was observed.

The absence of any substantial change in the αC-helix and A-loop regions was surprising, even more considering that in the case of the Q275A/P299Q/Q420E mutant, Q420 is the only residue that differs between the A-loop of Lck and Src. A further attempt was the substitution of the entire αC-helix of Src with the one of Abl. Unexpectedly, the chimera generated by mutating 14 residues determined the complete suppression of the sub-μs dynamics, not only of the αC-helix but of the Src structure

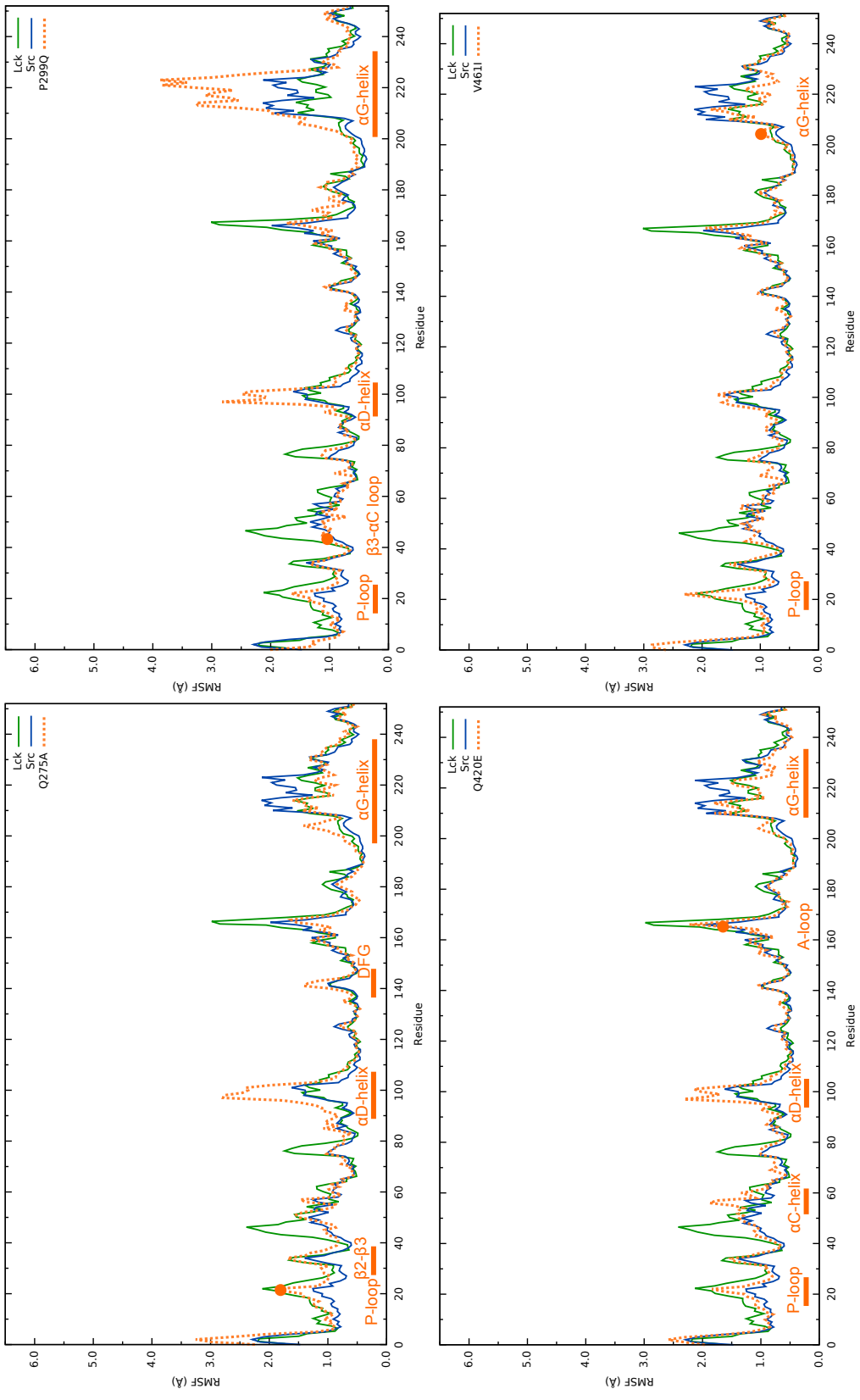

Figure 5.2: RMSF of the engineered single mutants of Src (Q275A, P299Q, Q420E, V461I). The RMSF profile of every mutant (dashed orange line) is plotted together with the two references profiles of Lck (green line, flexible) and Src WT (blue line, rigid). The orange dots identify the site of the mutation and the orange bars the regions of the structure whose fluctuation results affected by the mutation.

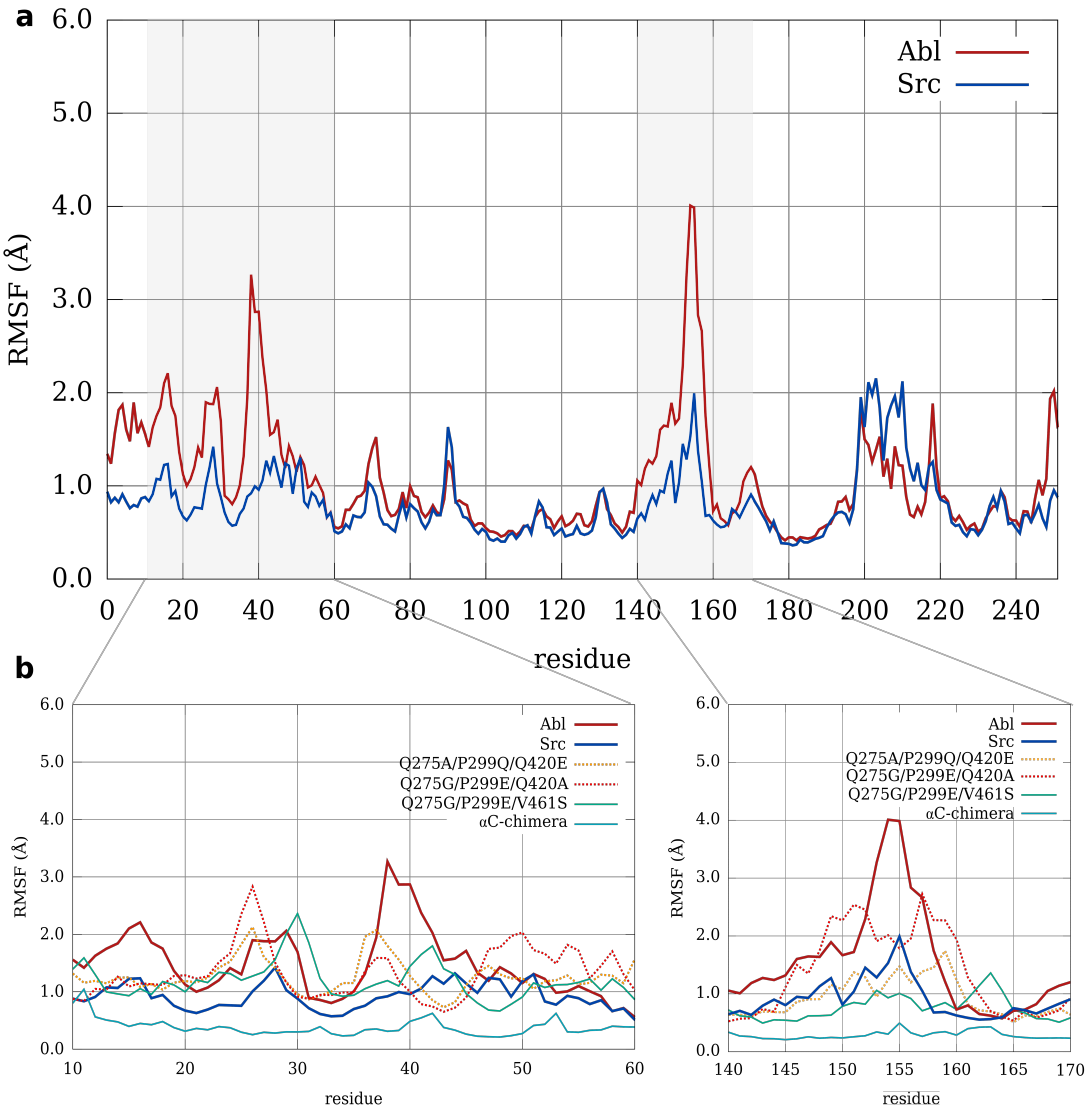

Figure 5.3: (a) RMSF profiles of Src (blue) and Abl (red). Grey panels identify the N-lobe and the A-loop region. (b) RMSF profiles of the engineered mutants of Src. In the two panels are reported the RMSF profiles of the N-lobe region (left) and A-loop region (right). Shades of red identify the more flexible mutants, shades of blue the more rigid. Dotted lines are used for clarity. Figure taken from Ref [93].

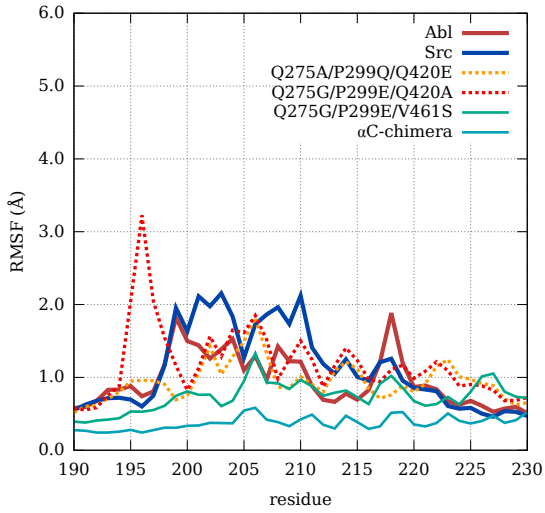

Figure 5.4: RMSF profile of the αG-helix of Src, Abl and the engineered mutants of Src. Shades of red identify the more flexible mutants, shades of blue the more rigid. Dotted lines are used for clarity. Figure taken from Ref [93].

in general, as shown in Figure 5.3 and 5.4.

In the attempt to flexibilise the A-loop, new mutations were designed in such a way to disrupt crucial contacts within the Src structure, focusing on those stabilising the αC-helix and the A-loop (see Figure 5.1).

Crystal structures of Src show the A-loop forming a helix turn, supposed to stabilise the loop in an open active-like conformation. The same behaviour was also reported in computational studies of EGFR[100, 214] and FGFR[103]. Therefore, mutations E415G and Y416G, disrupting the helix turn, were attempted (see Figure 5.1). Single and double mutants (E415G/Y416G) were tested for flexibility *in silico*, also a triple mutant bearing these two mutations together with the one on the αG-helix (V461S, see Figure 5.1). Again, a weak change in flexibility was achieved but only in the A-loop of the triple mutant (see Figure 5.5).

Subsequently, to increase the flexibility, this time of the αC-helix, a glycine was introduced (M302G, see Figure 5.1) to create a flexible point in between the helix and the preceding loop (β3-αC loop). The M302G mutant determines change in the overall Src dynamics very similar to the Q275G/P299E/Q420A mutant (Src to Abl), with the exception of the αG-helix region, not rigidified (see Figure 5.5).

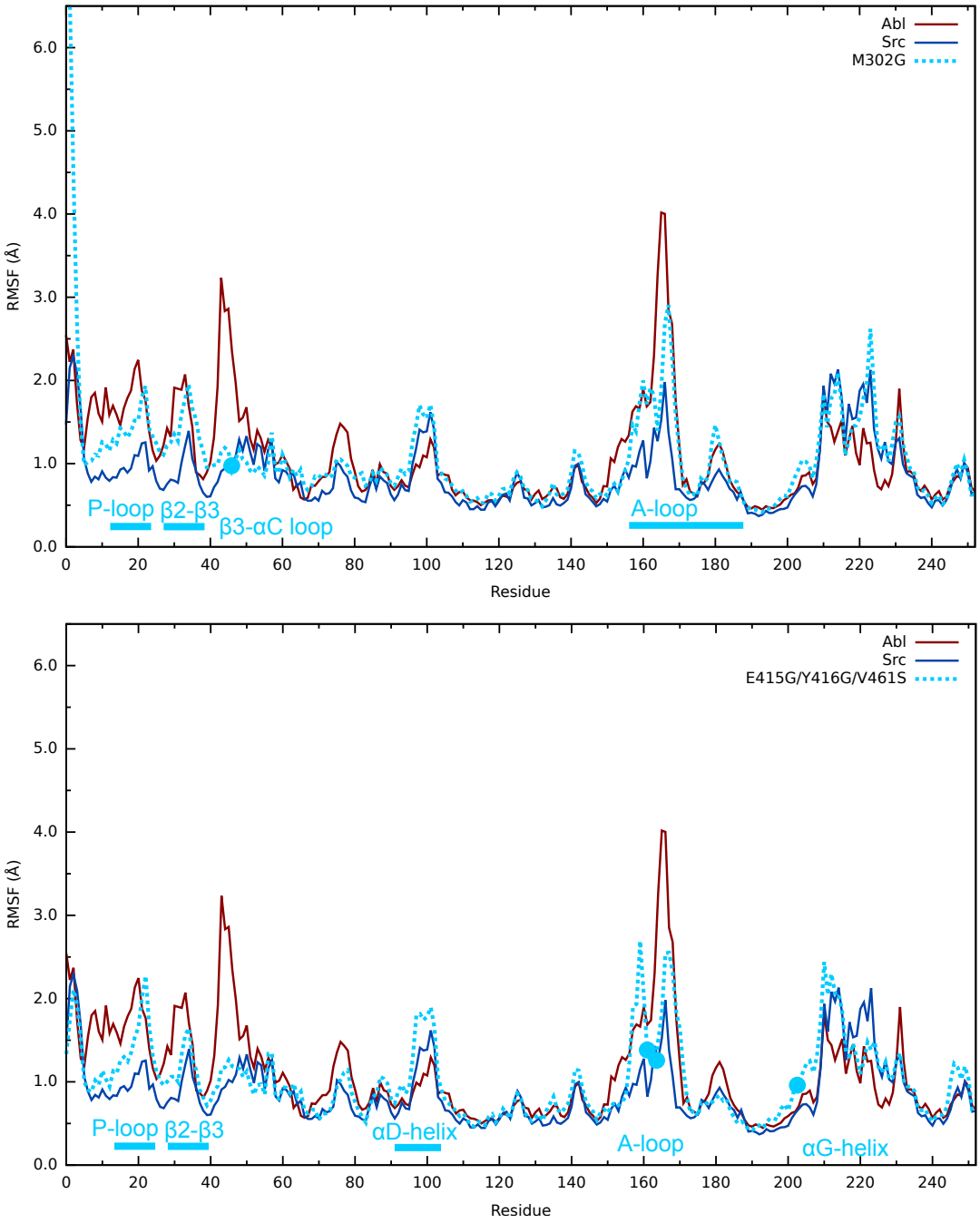

Figure 5.5: RMSF of the engineered mutants of Src: M302G and E415G/Y416G/V461S. The RMSF profile of the mutant (cyan dotted line) plotted with the two references profiles of Abl (red line, flexible) and Src WT (blue line, rigid). The dots in cyan identify the sites of the mutation and the bars the regions of the structure whose fluctuation results affected by the mutation.

5.3 Expression and purification of the engineered mutants of Src

The triple mutants, showing major differences in the sub-μs dynamics predicted *in silico*, were tested for imatinib affinity. The final candidates for expression and purification were: Q275A/P299Q/Q420E (Src to Lck), Q275G/P299E/Q420A (Src to Abl), Q275A/P299Q/V461S (Src to Abl, with mutation of the αG-helix) and finally the αC-chimera as a counterexample, given the suppressed dynamics. The mutants were expressed in *E. coli* cells, purified and their affinity towards imatinib tested by means of isothermal titration calorimetry (ITC) experiments.

The expression of the triple mutants and of the Src chimera resulted in a very extensive molecular biology study. The introduction of up to three mutations in the Src structure was sufficient to change completely the protein behaviour. In fact, it was impossible to express the mutants following the same protocol successfully used for the expression of Src WT. It followed the need to design an alternative protocol for the expression and also the purification of the mutants (for details see Section 2.2.3). In Figure 5.6 are shown the chromatograms of the two purification steps performed (affinity chromatography (image a) and size-exclusion chromatography (image b)). In image c, Figure 5.6, there are the gels showing the mutants obtained at the end of the purification process.

5.4 ITC experiments of the engineered mutants of Src

The experimentally measured K_d (see Figure 5.7) appears to correlate well with the RMSF values registered for the mutants (see Table 5.1, lower the K_d higher the affinity of imatinib for the mutant). It progressively decreases as the A-loop fluctuations increase and when the αG-helix is stabilised, as in mutant Q275G/P299E/V461S (see Figure 5.3 and 5.4). At the same time, the K_d of imatinib for the αC-helix chimera increases to14 μM (Src WT=12 μM), thus constituting a cross-validation for which a lower imatinib affinity is associated with the loss of sub-μs dynamics in the kinase.

Albeit, all Src mutants registered a decreased K_d (thus a higher sensitivity to imatinib), the change in the binding affinity of the drug was smaller than what one would have expected, assuming that the sub-μs dynamics in itself plays a major role in the

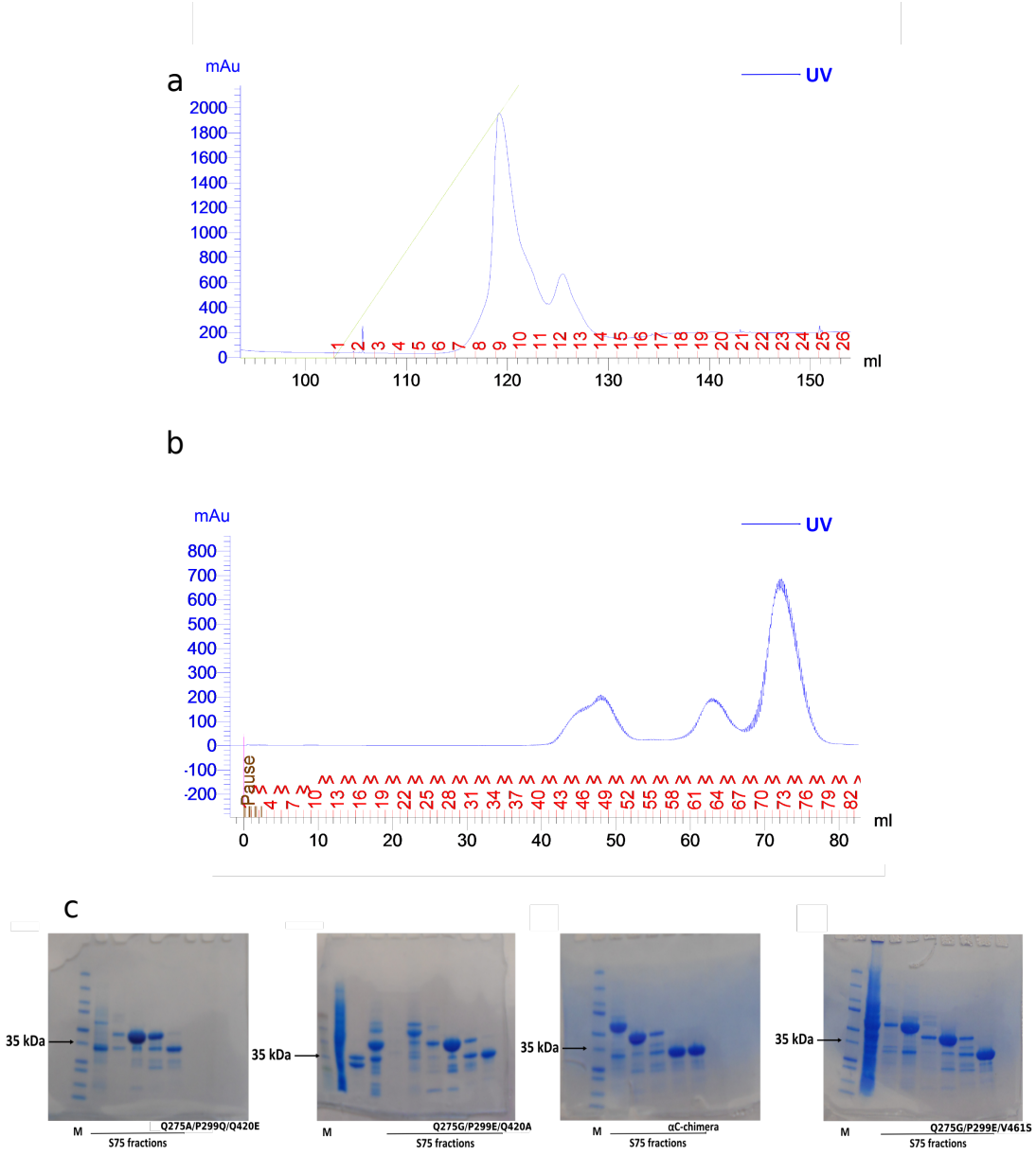

Figure 5.6: (a) HisTrap column chromatogram: the protein of interest is the more intense UV signal. (b) size-exclusion S75 column chromatogram: the protein of interest corresponds to the last signal. (c) Gels identifying the obtained mutants of Src after purification (around 35 kDa). Figure taken from Ref [93].

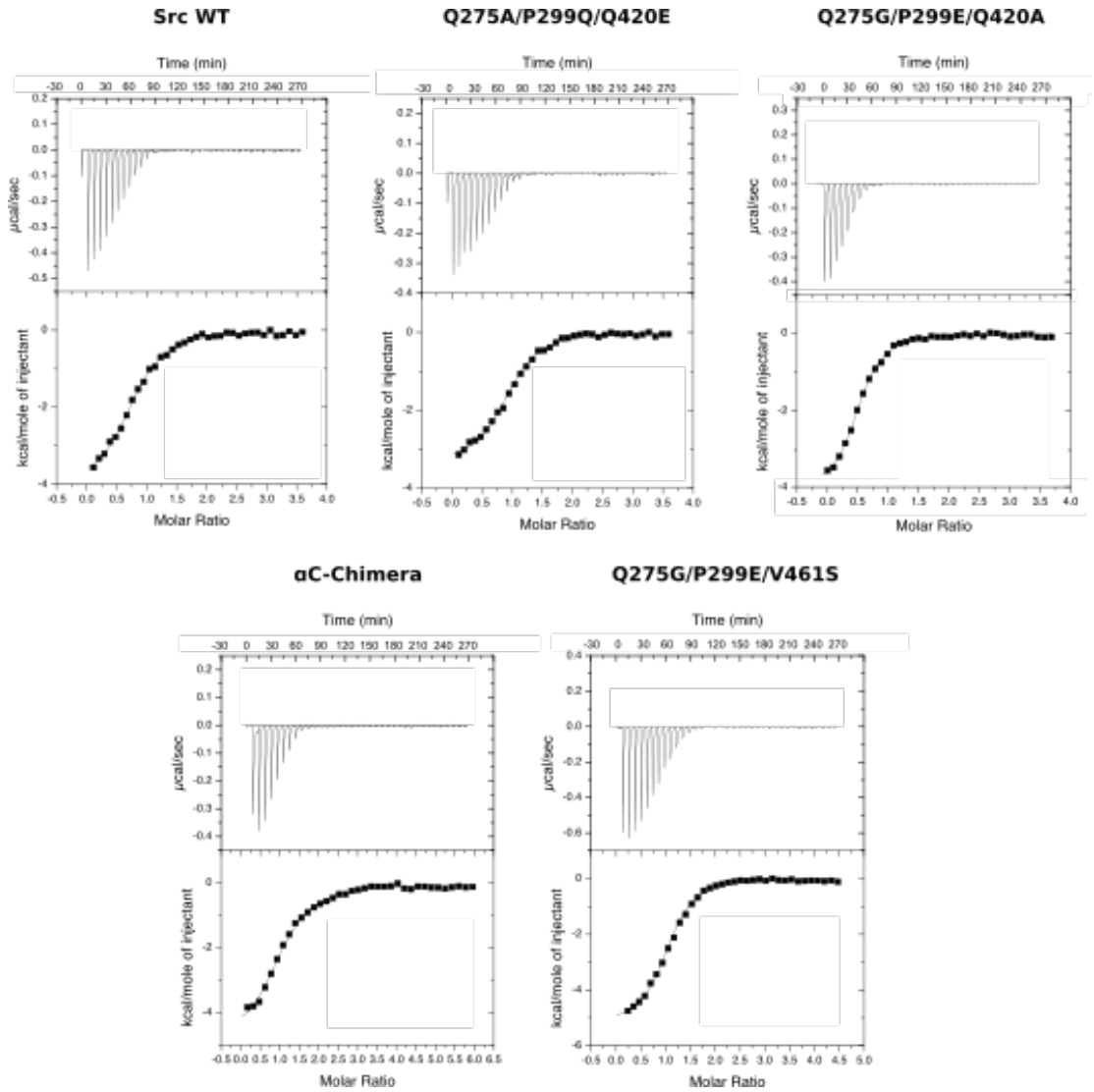

Figure 5.7: ITC curves resulting from the experiments performed and reporting the imatinib binding curves towards Src WT and the engineered mutants of Src. Figure taken from Ref [93].

Protein	A-loop RMSF (Å)	K_d (μM)
Abl WT	4.2	0.08 [20]
Q275G/P299E/V461S	1.4	6.2
Q275G/P299E/Q420A	2.5	7.3
Q275A/P299Q/Q420E	1.8	8.5
Src WT	2	12
αC chimera	0.3	14

Table 5.1: RMSF values of the A-loop region, averaged from MD calculations, and the retrieved K_d for Src WT and the engineered mutants of Src by means of ITC (data published in Ref [93]). Lower the K_d, higher the affinity of imatinib for the kinase. The K_d value of Abl was recovered from the work of Seeliger et al.[20].

binding. However, in contrast with previous studies in which similar K_d values were obtained[20] (see Figure 5.8), it is important to notice that all the mutated residues in the work presented in this thesis were chosen far from the binding site and not engaging in direct interactions with the drug.

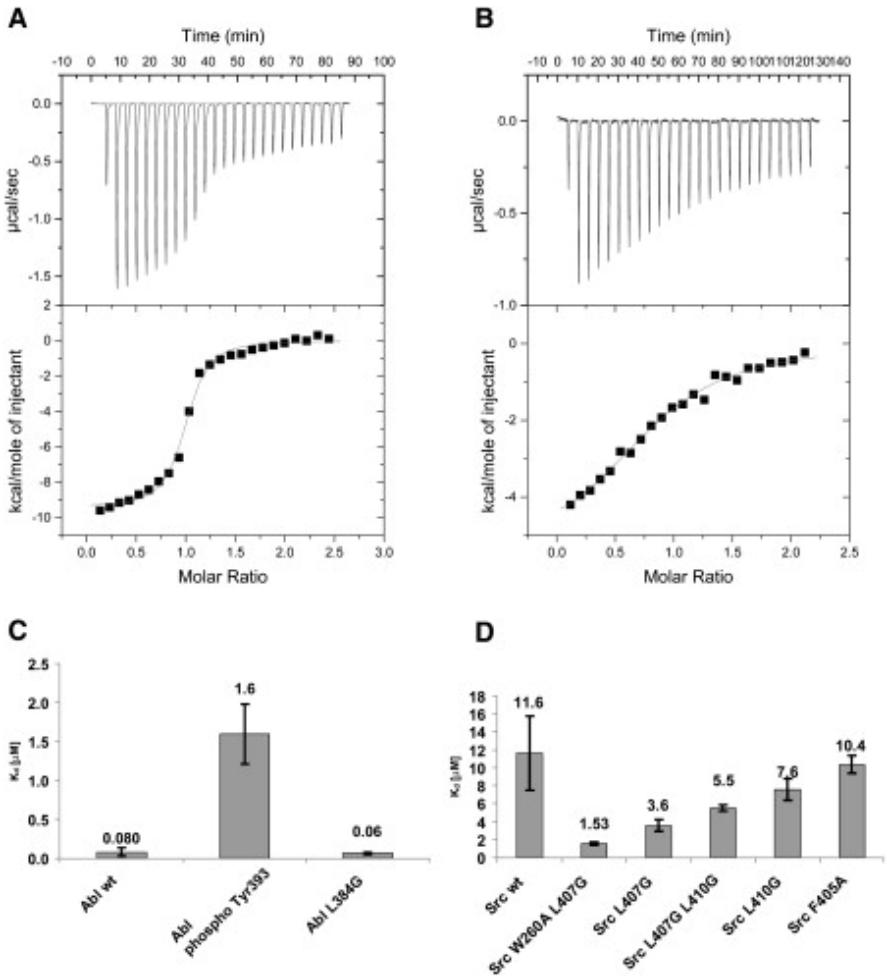

Figure 5.8: Figure taken from Ref [20]. ITC curves of imatinib binding to Abl (A) and Src (B). In (C) K_d values of imatinib obtained for Abl WT, Abl phosphorylated and Abl L384G mutant. In (D) K_d values of imatinib binding to Src WT and to some of the Src mutants described in the Ref [20]. Noticeable is the choice of the mutated residues, all in the immediate vicinity of the DFG motif (Asp404-Phe405)[20].

It is particularly interesting Q275A/P299Q/V461S, the αG-helix mutant, registering the highest affinity for imatinib. The RMSF shows a rigidification of the αG-helix, mimicking Abl WT, but not the flexibilisation of the A-loop. From the analysis of the MD trajectories, the αG-helix of Src WT was seen to adopt two main conformations, shown in Figure 5.9. On the contrary, the αG-helix of the mutant only adopts one (image A, in Figure 5.9), in which the helix is interacting with the A-loop stabilising both regions. Thus, the connection existing between these two regions in Src is not

allosteric, as thought at the beginning, but mainly due to formation and breakage of contacts. The rigidification of the αG-helix achieved in mutant Q275A/P299Q/V461S (mimicking Abl) is also responsible for the undesired stabilisation of the A-loop.

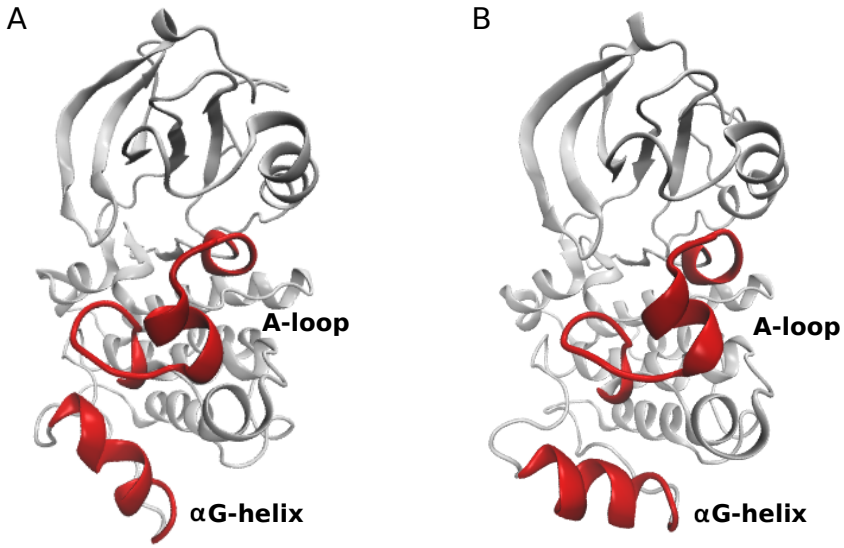

Figure 5.9: αG-helix conformations (A, B) sampled in the trajectory of Src WT. Figure taken from Ref [93].

5.5 Discussion

Based on the computational results, I performed a rational mutagenesis of the Src KD to make more flexible the rigid motifs (P-loop, β3-αC loop, αC-helix, A-loop) and vice-versa (αG-helix). The engineered mutants affect the sub-μs dynamics of the Src KD and, consequently, also imatinib K_d. This was a further evidence of a connection between sub-μs dynamics and imatinib affinity in TKs.

Furthermore, the mutagenesis study performed shed light on the complex allosteric network of Src summarised in Figure 5.10. It confirms the essential role of the αC-helix in the regulation of the KD, indeed its substitution in the chimera mutant determines a complete suppression of the dynamics. Also, the αG-helix and the P-loop regions (see Figure 5.10) show not only a strong cross-correlation between themselves, but also a high degree of connection to the overall kinase network, because always perturbed by other mutations. In particular, the αG-helix gains importance as strategic area in the overall KD regulation of Src, for the interactions it makes with the A-loop.

However, the flexibilisation of the A-loop was unsatisfactory for all mutants tested,

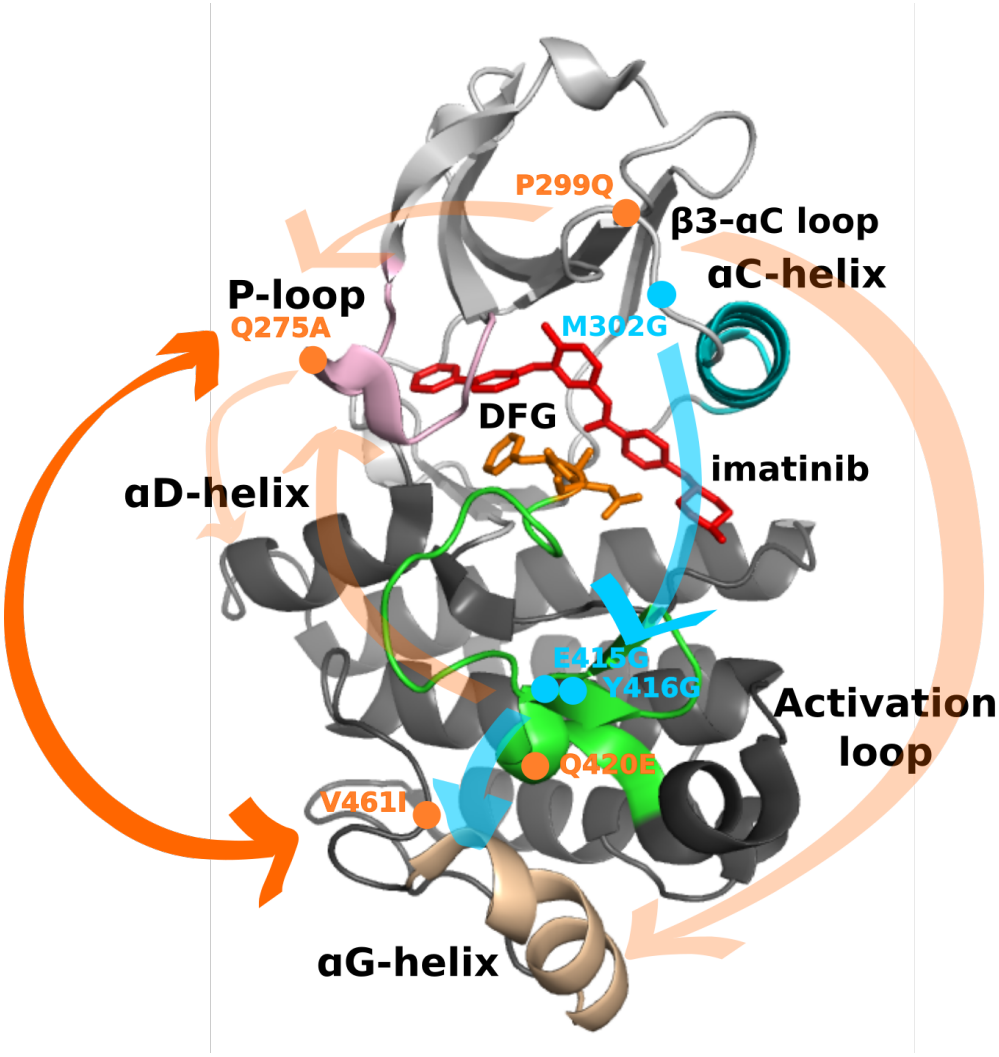

Figure 5.10: Structure highlighting the location of the engineered mutants of Src. The mutations affecting Src dynamics are coloured in orange, the ones affecting Src contacts and structure are coloured in cyan. Arrows are illustrating the major allosteric pathways identified by the mutagenesis study, coloured accordingly to mutant. The connection between P-loop and αG-helix is a feature of the Src structure.

and the subsequent change in the binding affinity of imatinib smaller than expected. The impossibility to effectively modify the dynamics of crucial areas involved in the conformational changes of Src, as the A-loop and in a lesser extend the αC-helix, could be explained considering a much more complex allosteric mechanism regulating the dynamics of these two regions[17, 41, 15]. For example, in a recent paper from Dölker and co-workers, a significant allosteric effect of the SH2 subunit acting on the αC-helix of Abl (refer to Figure 1.9 of the Introduction) was shown[41]. Similarly, in Src other subunits (SH2 and SH3) could be involved in the regulation of the KD.

What is clear is the presence of a complex interplay between allostery, sub-μs

dynamics and large conformational changes. In agreement with these considerations, it was also proposed in other works the presence of several concurring distributed factors, after that other extensive mutagenesis in Src and Abl failed to identify a single amino-acid responsible for the difference in imatinib activity[20]. However, the observation that sub-μs fluctuations are able to discriminate among kinases sensitive to imatinib might be used to predict the effect of selected mutations in cancer samples.

Chapter 6

Understanding of the molecular mechanism of imatinib resistance

In the previous chapters I reported evidence from computations and experiments showing that sub-μs dynamics is able to discriminate among different kinases, even when the sequences have more than 45% sequence identity. A special correlation was identified between the dynamics of regions involved in the conformational changes of the kinase and the activity of TKs inhibitors. The engineered mutants of Src shown that a gain in flexibility, even a small one, could enhance the affinity of imatinib to a kinase such as Src, which normally binds to imatinib with a very low affinity.

In this chapter, I will investigate the mechanisms of drug resistance in Abl. Most resistant mutants of Abl are extremely effective in impairing imatinib activity, even if they only harbour single point mutations. A well studied example is the "gatekeeper" mutant T315I[87, 94, 1], which was initially thought to block the binding due to a steric clash. Indeed, as many mutations are localised in the KD close to the catalytic pocket [88, 89], they could directly affect the binding. However, the mechanism of action of most mutations is still unknown as they are far from the binding site.

An exhaustive analysis was conducted by means of classical MD simulations and enhanced-sampling PTmetaD calculations, to investigate the dynamics of several resistant mutants of Abl. In agreement with the results reported in the previous chapters, the mutants analysed modify both the sub-μs dynamics and large conformational changes of the kinase, confirming once more the importance of dynamics in imatinib activity.

6.1 Selection of the Abl resistant mutants included in the study

To assess in detail how the dynamics is affected when resistance-inducing mutations arise in the kinase, four mutants of Abl were selected. The candidates were chosen among the clinical mutations found in chronic myeloid leukemia patients and collected in the works of Azam *et al.*[88, 89], having an unknown mechanism of resistance and a big impact on the IC_{50} of imatinib. The final mutants were: G250E, a single point

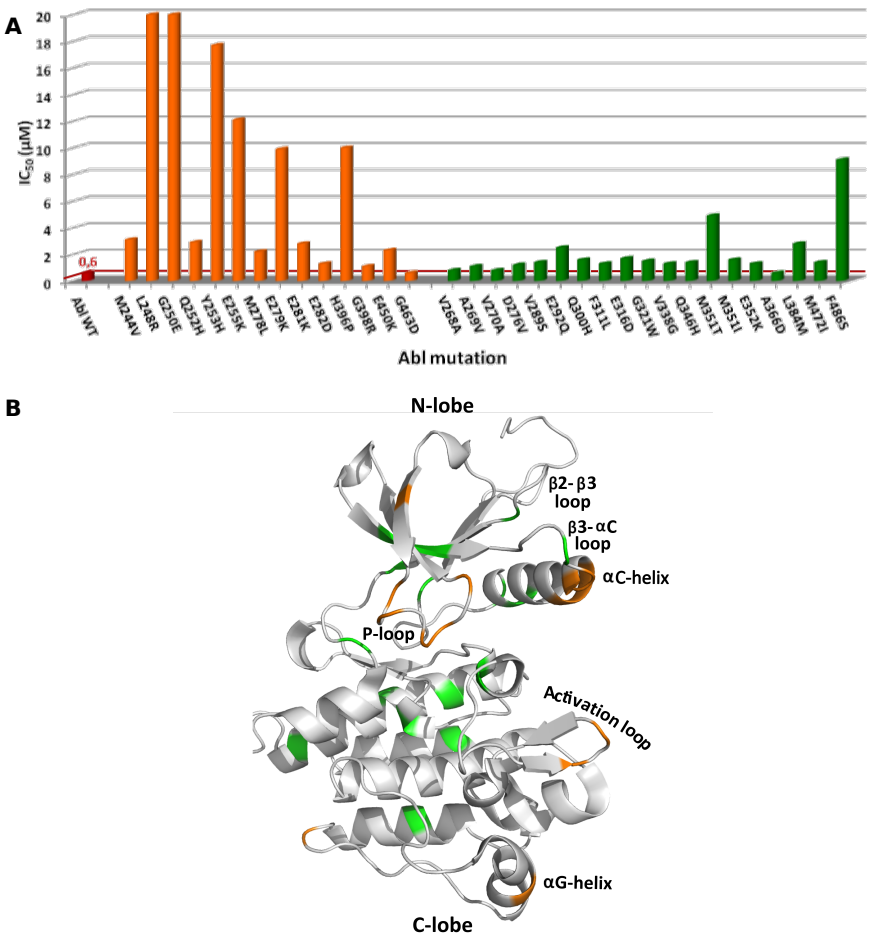

Figure 6.1: (A) IC_{50} values of imatinib corresponding to the resistant mutants of Abl collected in the work of Azam *et al.*[88], compared to the IC_{50} of Abl (in red). The mutations are divided in those lying in flexible regions of the Abl KD (in orange) and those lying in more constrained areas (in green), also directly highlighted on the Abl structure (B). As it could be noticed the mutations on flexible regions (P-loop, A-loop, αC-helix, αG-helix) have a more drastic impact on the IC_{50} (higher is the value, lower is the inhibitory effect of imatinib towards Abl).

mutation located on the P-loop (IC_{50} = 20 μM), E279K, mutant of the β3-αC loop (IC_{50} = 9.9 μM), H396P, mutant of the A-loop (IC_{50} = 9 μM) and E450K, mutant of

the αG-helix (IC_{50} = 2.3 μM).

Investigating the pathological mutations of Abl[88], it was noticed that the ones that mostly affect the IC_{50} of the drug, were prevalently located in regions of the structure recognised as flexible in the previously performed RMSF analyses (see Chapters 4, 5 and Figure 6.1). Being mutational hotspots, is *per se* a proof of the functional role played by flexible regions with regards to imatinib activity.

For completeness, the "gatekeeper" mutant of Abl, T315I ($IC_{50} > 20$ μM)[88] was added to the pool of mutants under study. Diverse mechanisms of action have been proposed for this mutant: a "direct" one, with the elimination of an hydrogen bond crucial for the high-affinity binding of imatinib[91, 61], or to impede the induced fit mechanism proposed for Abl[79, 80], determined by the adoption of a kinked conformation of the P-loop, or even the modulation of the activation mechanism *via* conformational selection, as shown for others "gatekeeper" mutants in homologous kinases[100, 101, 102, 103].

6.2 Classical MD simulations and RMSF analysis of the resistant mutants of Abl

MD simulations and RMSF analyses were performed for all the resistant mutants of Abl (for details see section 2.1.4). The results show a considerable change in the flexibility, with a clear tendency of all resistant mutants to adopt a more rigid "Src-like" profile. On average, the difference in RMSF between Abl WT and the mutants is of about 1.5 Å (see Figure 6.2). The main regions to be affected are again the N-lobe, the A-loop, and, to a lesser extent, the αG-helix.

In the G250E mutant, the N-lobe loops are largely affected (P-loop, β2-β3 loop and β3-αC loop) together with the A-loop, with a RMSF decrease of about 1 Å and almost 3 Å, respectively. The E279K and H396P mutants show similar RMSF profiles. In both cases, there is an important decrease in flexibility with respect to the WT in the β3-αC loop and, particularly, in the A-loop (of about 2 Å). The allosteric communication, between the β3-αC loop and the A-loop in Abl, results in both loops being affected even when only one of the two bears the mutation. The β3-αC loop is known to play a role in the activation of Abl, *via* allosteric regulation by the SH2 domain[41, 213],

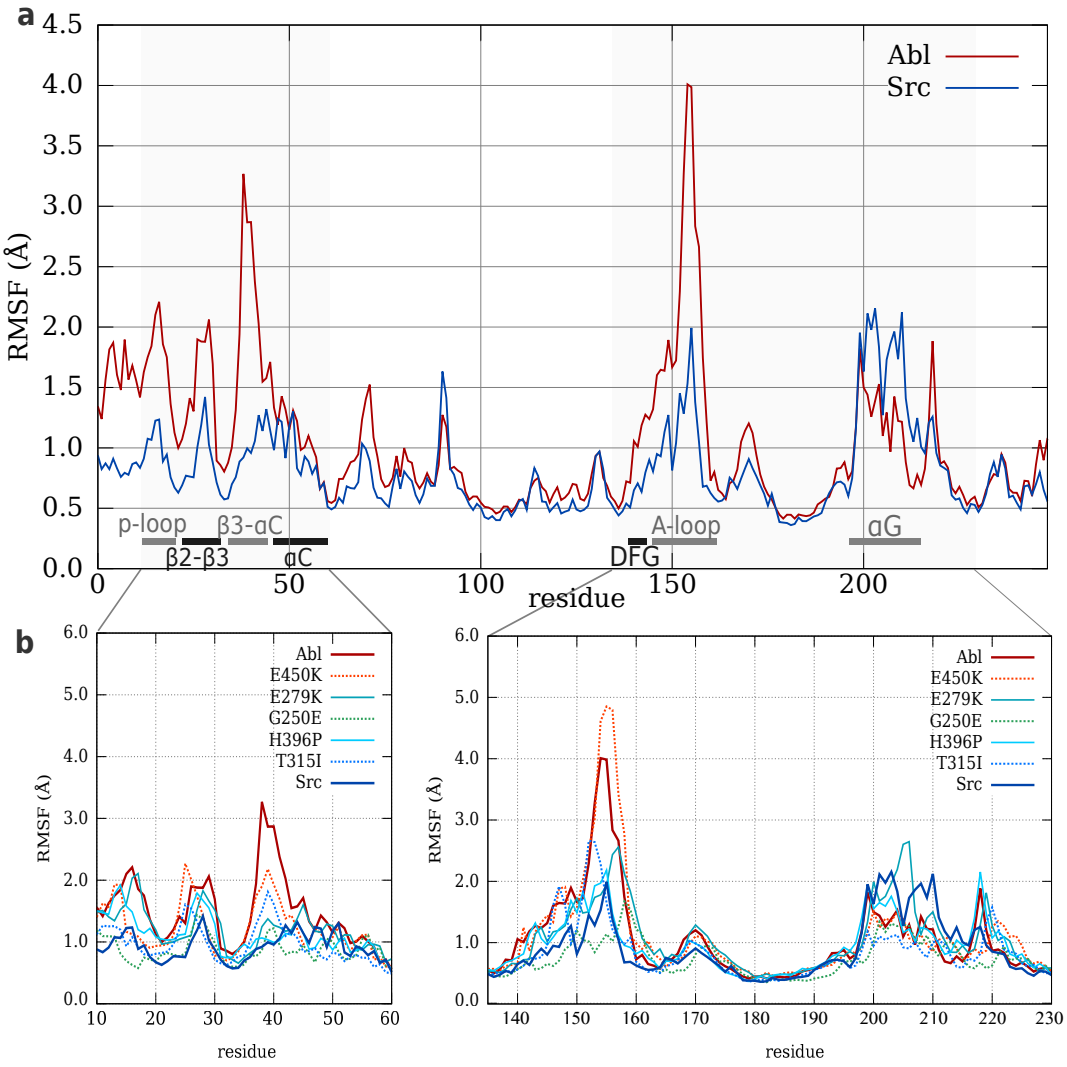

Figure 6.2: (a) RMSF profiles of Src (blue) and Abl (red). Grey panels identify the N-lobe and the A-loop with αG-helix region. (b) RMSF profiles of the resistant mutants of Abl. In the two panels are reported the RMSF profiles of the N-lobe region (left) and A-loop with αG-helix region (right), being the areas whose flexibility varies most among the pool of TKs. Shades of red identify the flexible mutants, shades of blue the more rigid ones. Dotted lines have been used for clarity. Figure taken from Ref [93].

so its connection with the A-loop becomes even more relevant. A small but significant increase in RMSF can also be observed for the αG-helix, especially in the E279K mutant. Only minor changes, were instead registered for the αG-helix mutation (E450K), where the only affected region is the β3-αC loop.

Apparently, the decrease in flexibility, especially in the N-lobe and A-loop regions, although far from the binding site, is reflected in a loss of imatinib activity, in line with what was observed for the pool of homologous TKs, previously discussed (see Section 4.3).

6.3 Principal component analysis

To investigate if the Abl dominant motions were also affected by the mutations, PCA analysis was performed for the resistant mutants of Abl. As previously described in Section 4.4, the first two eigenvectors of Abl wild-type (WT) were used. The first eigenvector describes the movement of the N-lobe (hinge motion), while the second eigenvector describes mainly the lobe twist (refer to Figure 4.4 of Chapter 4).

A change can be appreciate in particular for G250E, the P-loop mutant, in the subspace it explores (see Figure 6.3). Most affected are the movements of the N-lobe, represented by the first eigenvector, which are drastically reduced and shifted towards a different region of the conformational space compared to the wild-type. The A-loop movement, corresponding to the second eigenvector, shows a more restricted ensemble of conformations. For E279K, H396P and E450K, the space explored overlaps with the one of Abl WT, although reduced along both directions. For the gatekeeper mutant, T315I, the space explored is substantially reduced along eigenvector 1 while, almost identical to the WT along eigenvector 2.

On the whole, the analyses of the sub-μs dynamics show a connection between the A-loop fluctuations and imatinib activity. Not only homologous TKs with lower RMSF seem to be less sensitive to imatinib (see Chapter 4), but also the pathological drug-resistant mutations of Abl induce a similar shift in the sub-μs dynamics.

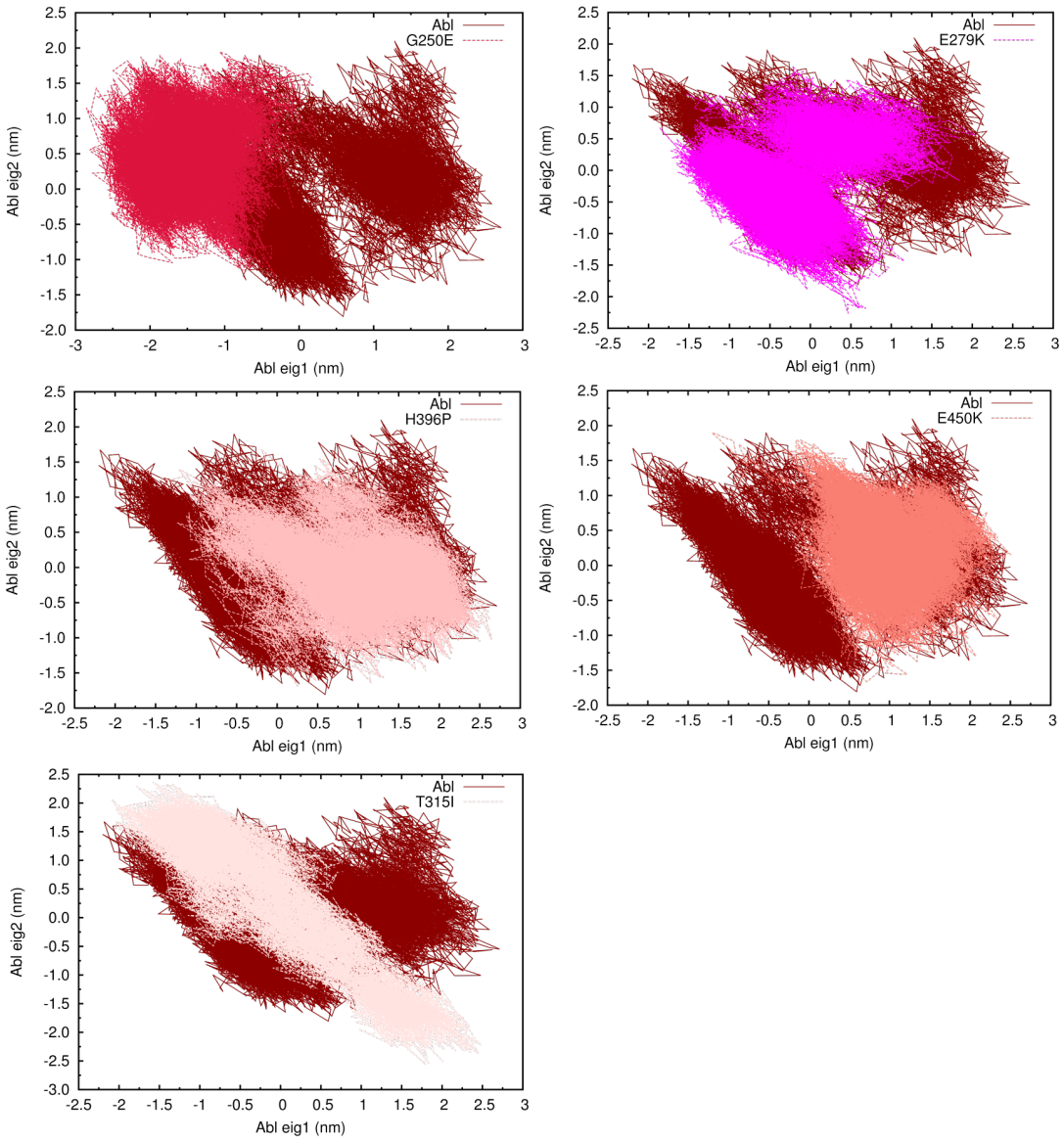

Figure 6.3: PCA analysis. Comparison of the conformational space explored by the resistant mutants of Abl with respect to Abl WT.

6.4 PTmetaD-WTE calculation and free-energy surface reconstruction

The observed changes in the sub-μs dynamics may probably be a proxy for a more global entropic impediment of the inactive DFG-out conformation itself, or in other words, a shift towards the kinase active conformation. Recently, it was demonstrated that over-activation of kinases, in the presence of oncogenic mutants, is indeed connected with a population shift towards the active DFG-in conformation, accompanied by some degrees of rigidification[100, 101, 102, 103].

To further investigate how the conformational landscape of Abl WT changes due to mutations, massive multiple-replica PT-MetaD simulations were performed to compute fully converged conformational free energy landscapes (for details see Section 2.1.6 and Figures 6.5 and 6.6), associated with the four drug-resistant mutants (G250E, E279K, H396P, E450K) and the "gatekeeper" mutant T315I (see final FESs in Figure 6.4). The same set of CVs used in the PTmetaD simulations of Src and Abl WT (see Chapter 3) were considered. As expected, for all the mutants the global minimum of the FES corresponds to the DFG-in conformation (see Figure 6.4). The FESs also exhibit a second minimum corresponding to the DFG-out, much more pronounced in the case of G250E and T315I, while not always well-defined in the others (e.g. E279K and E450K).

All resistant mutants show a significantly altered landscape, in particular H396P, E279K and E450K, remarkably similar to the one of Src. The G250E and T315I mutants, show a ΔG and a barrier very closed to Abl WT(4.0 ± 0.5 kcal/mol and barrier 6.0 ± 0.5 kcal/mol). However, the disappearance of the third local minimum, present in Abl WT surface and corresponding to the intermediate Asp-down conformation, confirms clearly that the DFG flip mechanism has been significantly altered by the mutations. The DFG-in minimum, in T315I, is particularly broad and much more populated than the DFG-out. While, the DFG-out conformation of G250E has a peculiar and never reported "out-out" conformation, adopted by both Asp404 and Phe405 (see Figure 6.7). The A-loop reaches the fully open conformation, accompanied by a loss of structural integrity in both the αC and αG helices (see Figure 6.7). Even if energetically the DFG-out does not seem to be affected, structurally the unfolding events and the inappropriate DFG-out conformation could account for changes in the binding mechanism of imatinib. The same out-out position of Asp404 and Phe405 has been observed also in the mutant E450K (Figure 6.7), that has a less stable DFG-out conformation, ΔG of 6.0 ± 0.5 kcal/mol, but an energy barrier for the DFG flip equal to Abl WT (around 6.0 ± 0.5 kcal/mol).

In the other two mutants, H396P and E279K, the DFG flip results to be even more strongly hindered with a ΔG of 7.0 ± 0.5 kcal/mol and 8.0 ± 0.5 kcal/mol, respectively, and an energy barrier of 9.0 ± 0.5 kcal/mol closer to the one registered for Src (11.0 ± 0.5 kcal/mol). Thus, not only the energy barrier is increased but the DFG-out minimum

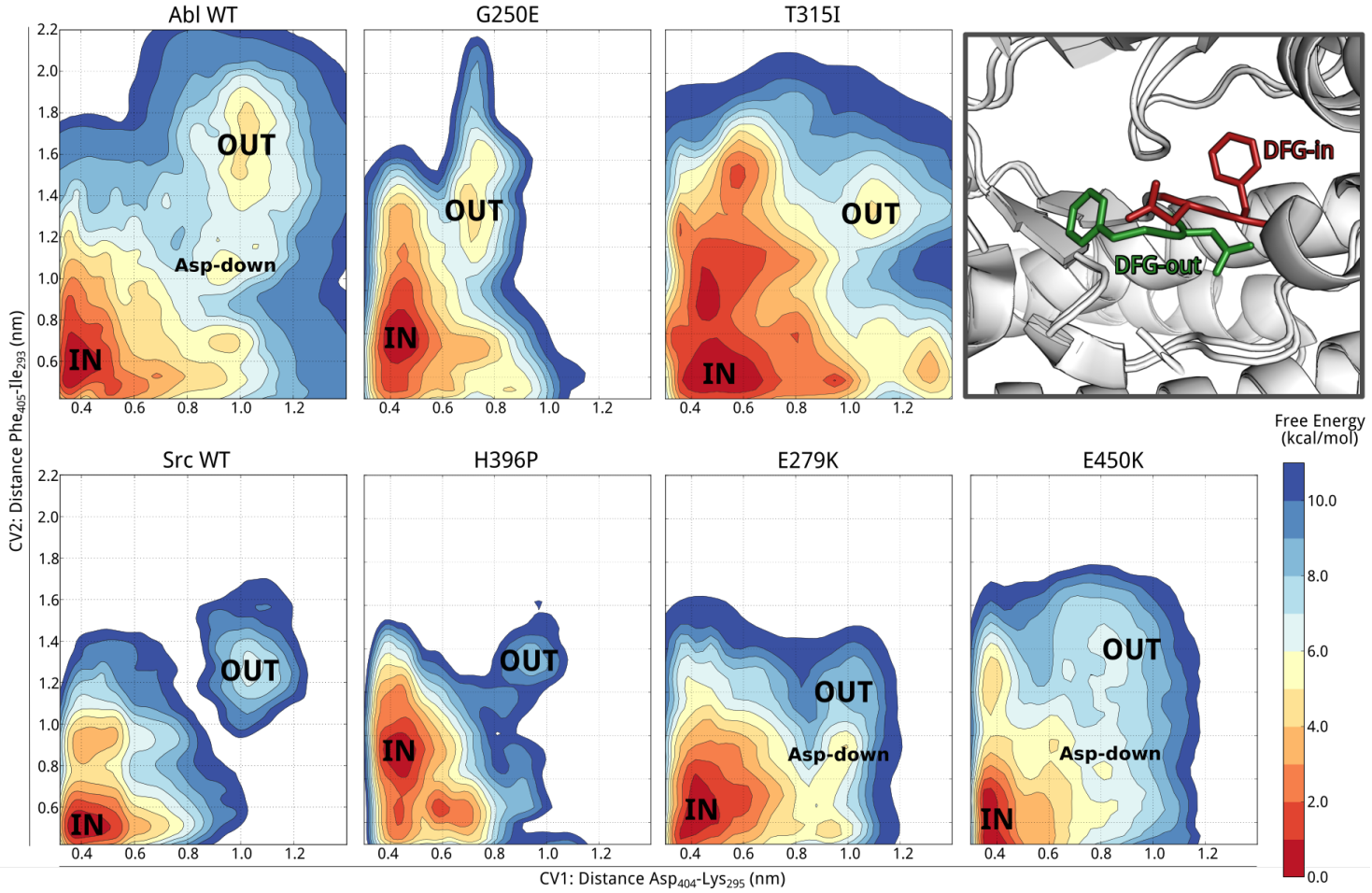

Figure 6.4: Reconstructed FESs for Abl WT, Src and the resistant mutants of Abl in function of CV1 and CV2. The DFG-in minimum is identified by the "IN" label, the DFG-out minimum by the "OUT" label and the intermediate Asp-down conformation, when present, by the label "Asp-down". The FESs of Src and Abl WT were taken from Chapter 3 and readapted. The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. Figure taken from Ref [93].

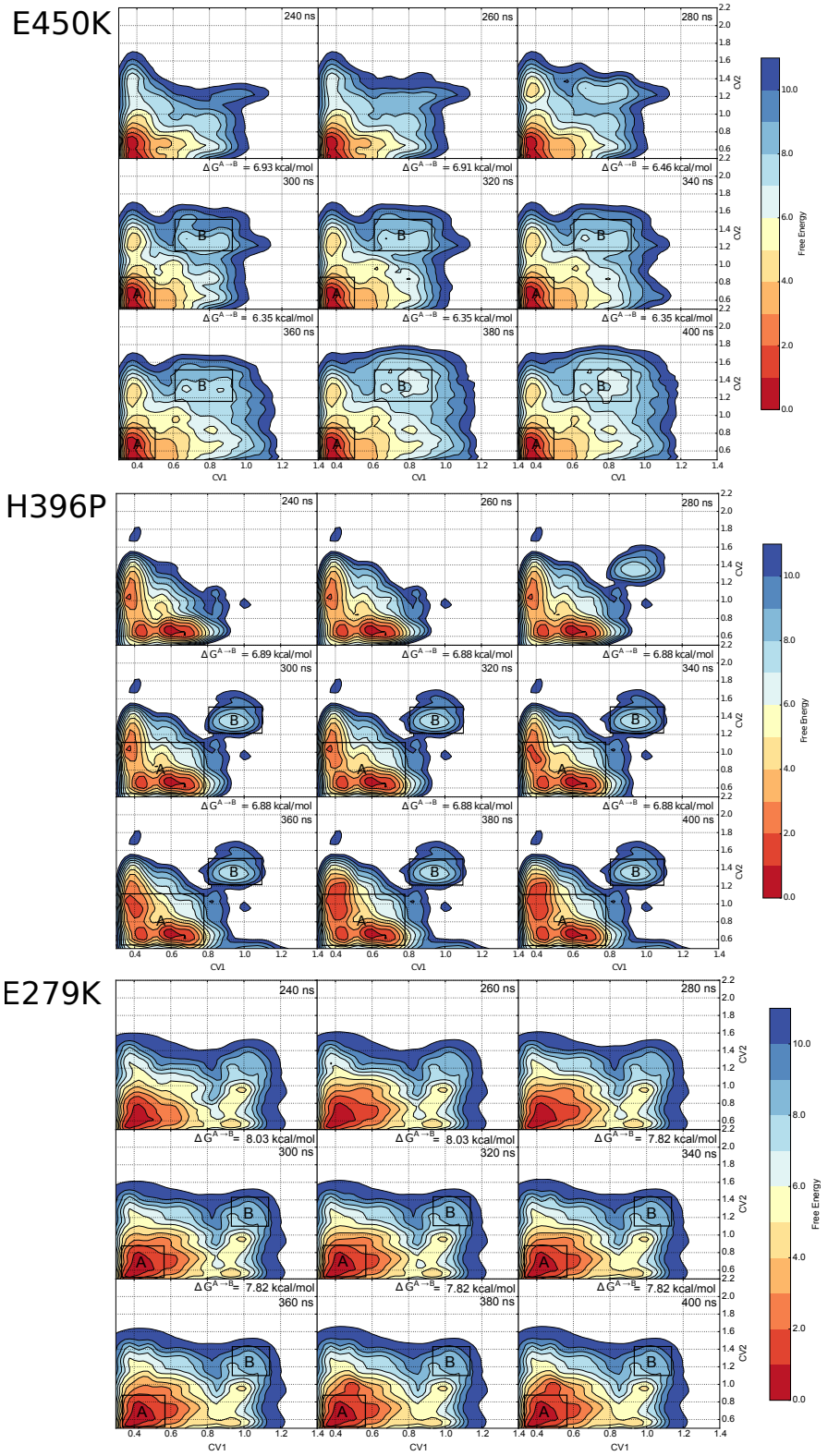

Figure 6.5: 2D projections showing the convergence of the free-energy of the resistant mutants of Abl (E450K,H396P and E279K). The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. Figure taken from Ref [93].

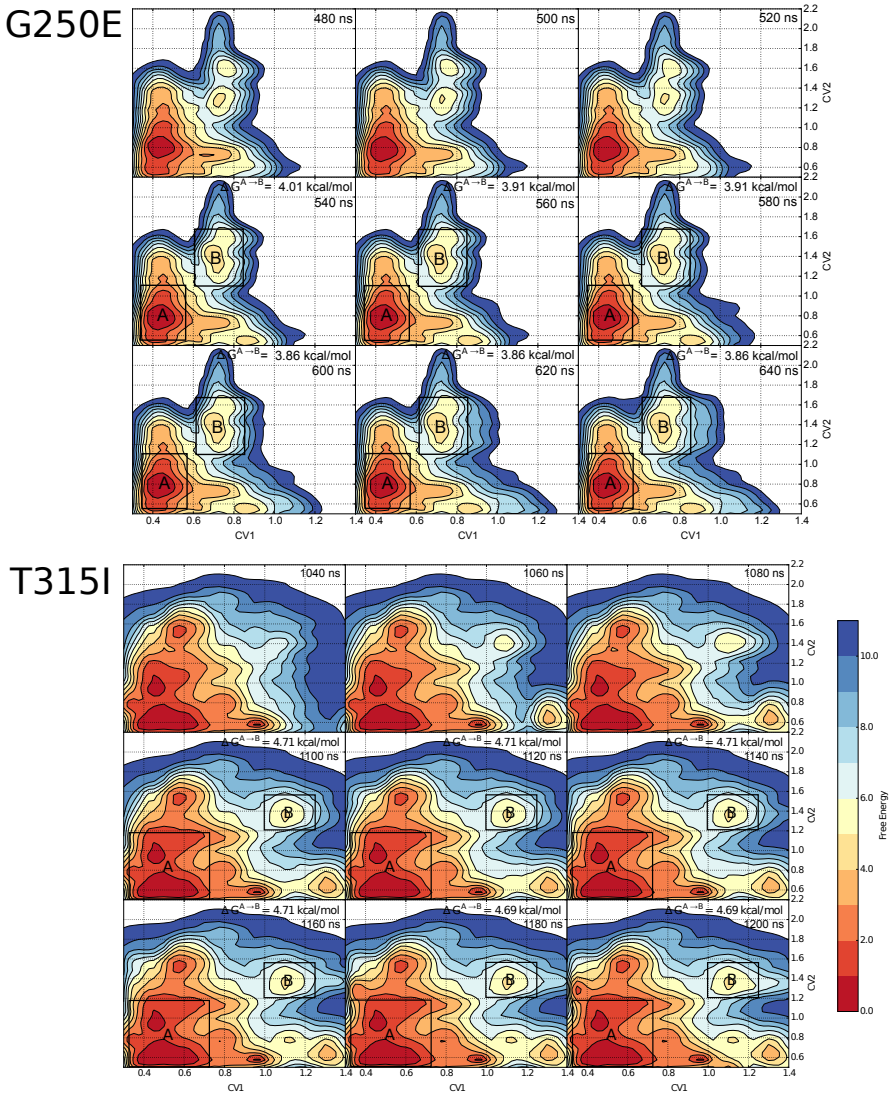

Figure 6.6: 2D projections showing the convergence of the free-energy of the resistant mutants of Abl (G250E and T315I). The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. Figure taken from Ref [93].

is less stable than in Src (6.0 $\pm$ 0.5 kcal/mol). For the E279K and E450K mutants, the DFG-out conformation is not even an energy minimum at all. Instead, the Asp-down conformation seems to be more stable and to be a secondary minimum for E279K and, in a lesser extend also for E450K. The DFG-out conformation in H396P is characterised by the complete unfolding of the αC and αG helices (see Figure 6.7). The loss of integrity of these two structural elements, critical for kinase functionality, could be a consequence of the higher rigidity of the structure, leading to the A-loop not being able to rearrange properly in the flipped conformation. Moreover, the A-loop in the DFG-out conformations of all mutants tends to be more helical, forming a second helix

turn, characteristic of the open A-loop of Src WT[214, 215, 103] (see Figure 6.8). The residues involved in the helix turn, from Asp391 to Ala395 (Abl 2G1T.pdb numbering), in the case of H396P are situated right before the mutation.

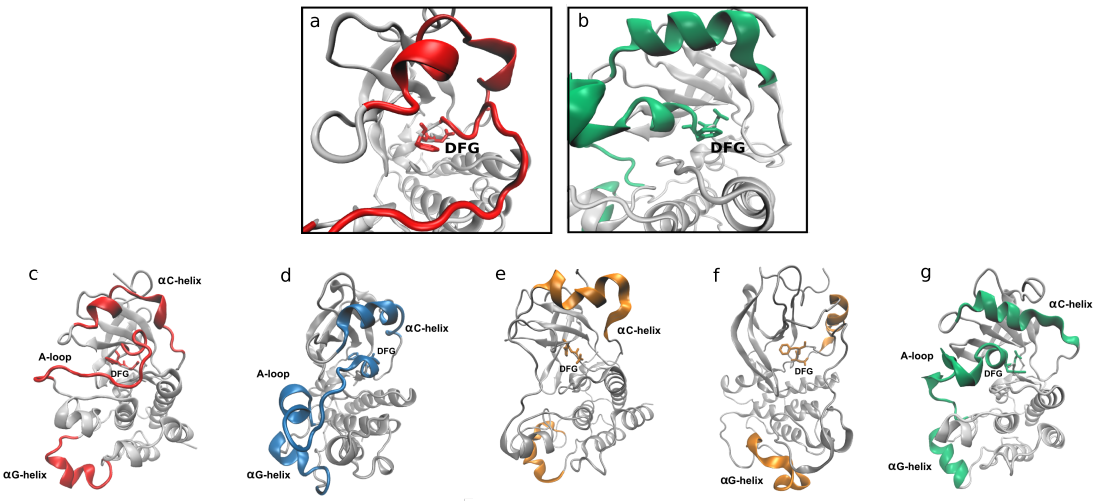

Figure 6.7: DFG out-out conformation identified in the DFG-out minima of G250E (a) and E450K (b). Unfolding of the αC and αG helices in G250E (c), E279K (d), H396P (e, f) and E450K (g). Figure taken from Ref [93].

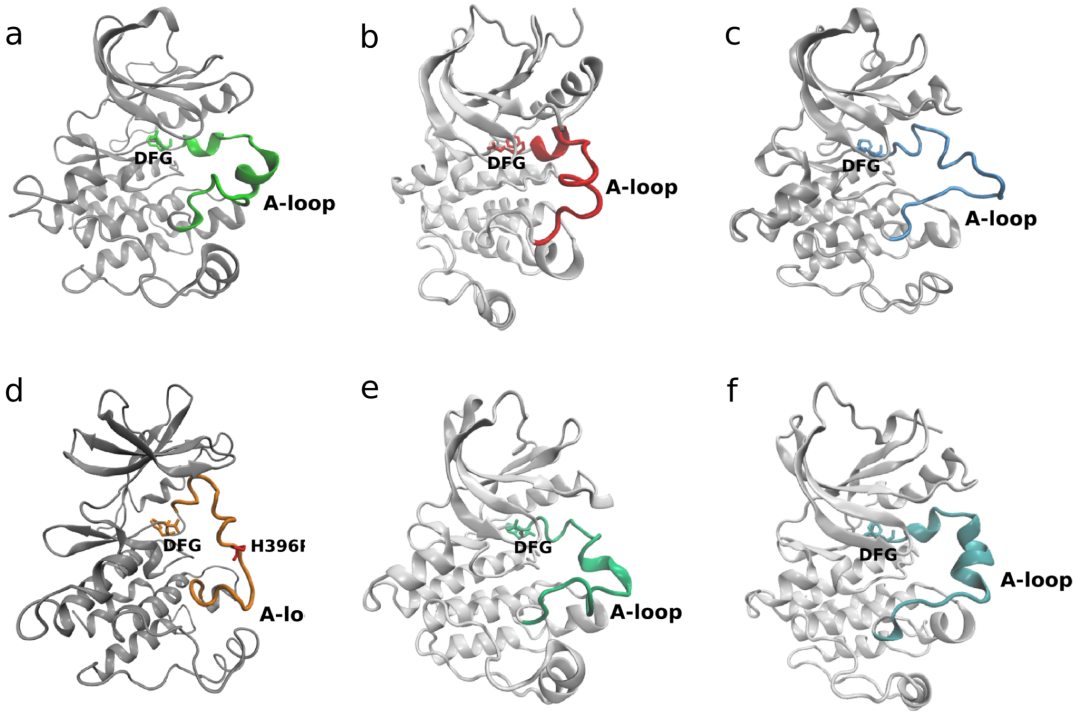

Figure 6.8: Structures of Src WT (a), G250E (b), E279K (c), H396P (d), E450K (e) and T315I (f) showing the formation of a helix turn in the A-loop of the resistant mutants of Abl, resembling Src. In sticks is shown the DFG motif and in H396P mutant in red the site of the mutation, located exactly after the helix turn. Figure taken from Ref [93].

Remarkably, all resistant mutants explore a reduced conformational landscape than Abl WT along both CVs, as observed for Src. In particular they never explore CV2 values above 1.7 nm. An effect which appears to be more significant in the case of the DFG-out conformation and may explain why the RMSF profiles are predictive of the drug activity.

6.5 Estimating the entropic and enthalpic contribution to the ΔG_{in-out}

The entropic and enthalpic contribution to ΔG_{in-out} in function of the temperature were estimated from the replicas of the PTmetaD simulations of Src, Abl WT and the resistant mutants. Figure 6.9 shows the linear regressions obtained for each simulation and in Table 6.1 are collected the corresponding entropic and enthalpic contributions retrieved.

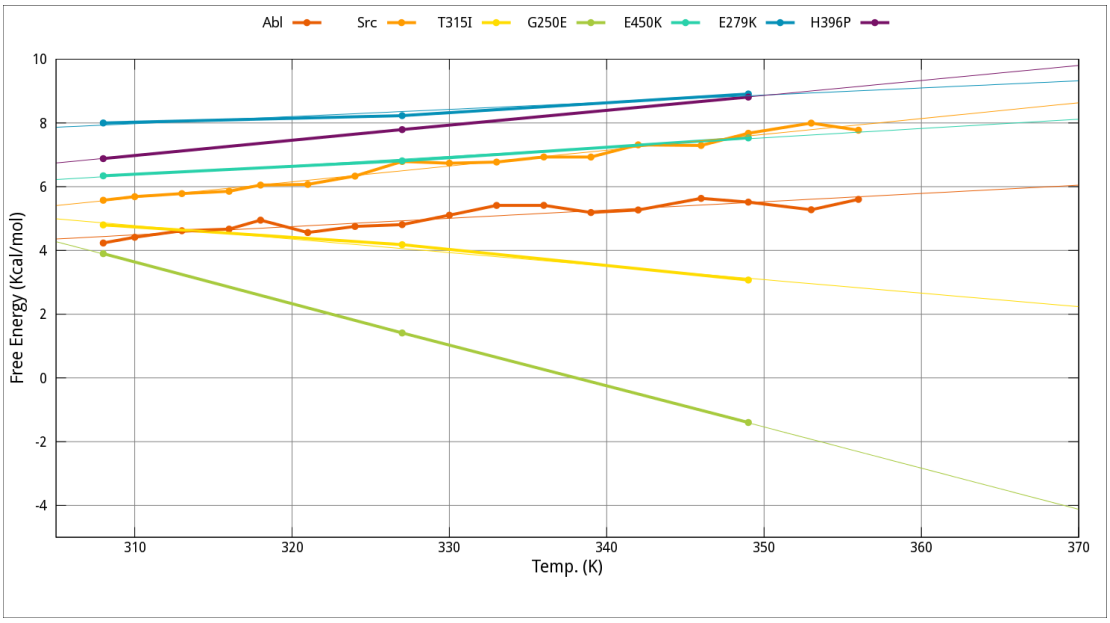

Figure 6.9: Linear regressions of Abl WT, Src and the resistant mutants of Abl. Each point of the regression corresponds to the ΔG_{in-out} calculated for each replica of the system at different temperature. This was repeated for the PTmetaD simulations performed. Figure taken from Ref [93].

The destabilisation of the kinase inactive DFG-out conformation for some of the resistant mutants is due to an entropic loss in the passage from DFG-in to DFG-out (see Table 6.1). The only exceptions are G250E and T315I mutants, for which the transition results to be entropically favoured (Table 6.1). Thus, classical MD simulations of

the DFG-out conformation of Abl WT, Src and the resistant mutants were performed (see details in Section 2.1.4) to check for the sub-μs dynamics also of the druggable conformation. Effectively, the RMSF analysis performed confirms that just for G250E and T315I, the DFG-out conformation experiences a slight increase in the fluctuations of the A-loop (see Figure 6.10).

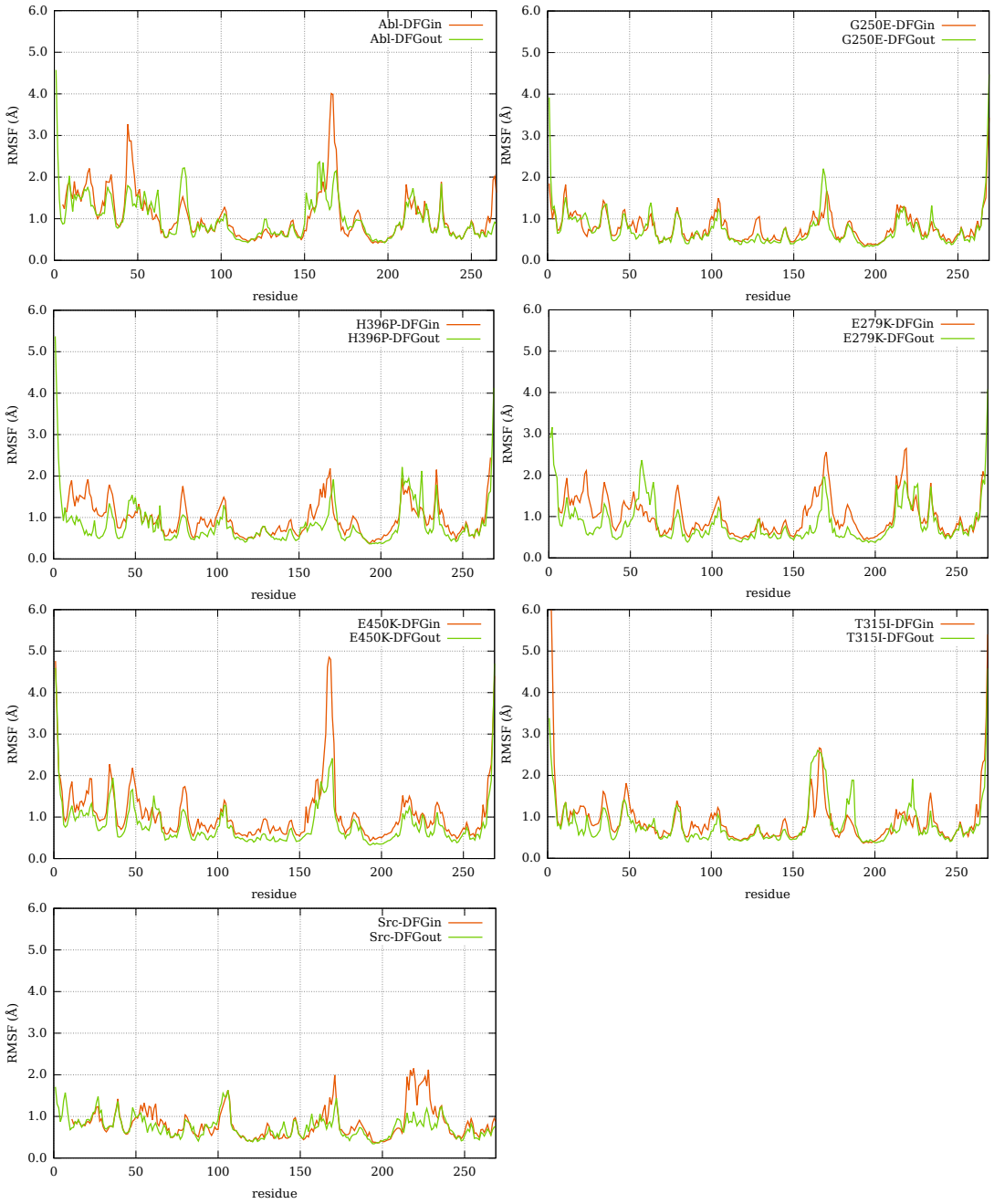

Figure 6.10: RMSF profiles of the DFG-in (in orange) and DFG-out (in green) conformations for Abl WT, Src and the reisstant mutants of Abl. Figure taken from Ref [93]

The destabilisation of the kinase inactive conformation caused by some ima-

tinib resistant mutants is in striking agreement with the published works on oncogenic mutants[98, 100], confirming that the destabilisation of the DFG-out drives a shift towards the DFG-in in both oncogenic and imatinib resistant mutations.

TK	ΔH (kcal/mol)	$T\Delta$S (kcal/mol)
Abl	-3.56	-8.01
E450K	-2.66	-8.96
T315I	17.93	13.06
H396P	-7.61	-14.51
E279K	1.02	-6.90
G250E	43.69	39.79
Src	-9.73	-15.28

Table 6.1: Entropic and enthalpic contributions to the ΔG_{in-out} in Abl WT, Src and all the resistant mutants of Abl. The values have been obtained from the linear regressions of the ΔG_{in-out} as a function of temperature shown in Figure 6.9 with equation $\Delta G=\Delta H-T\Delta S$. Figure taken from Ref [93].

6.6 Free energy reweighting

Still difficult to characterise is the role played by the mutations in G250E and T315I mutants. In literature it is mentioned that oncogenic mutants are able to cause hyperactivation of the kinase, as confirmed by a common mechanism of resistance that consists in locking the kinase in its biologically active conformation, ready to bind ATP and to phosphorylate substrates[102, 103]. Similarly, the previous calculations suggest that imatinib resistant mutations could not only destabilise the druggable, DFG-out conformation of Abl WT thus impairing imatinib binding, but probably also shift the A-loop towards more open 'active like' structures. It has been shown that the "active-like" conformation of TKs, with the A-loop open but not phosphorylated, is just marginally stable and can be stabilised by ATP binding[73, 100, 75].

To further investigate this aspect, I analysed how the mutations alter the open-closed equilibrium of the A-loop. The reweighting of the PTmetaD simulations was performed (see Section 2.1.8) in order to include two additional collective variables, able to describe the opening of the A-loop (see Figure 6.11).

In agreement with recent calculations on Src WT[73], the open conformation of the A-loop in both Abl and Src WT (CV1$<$6) are highly disfavoured (see Figure 6.11).

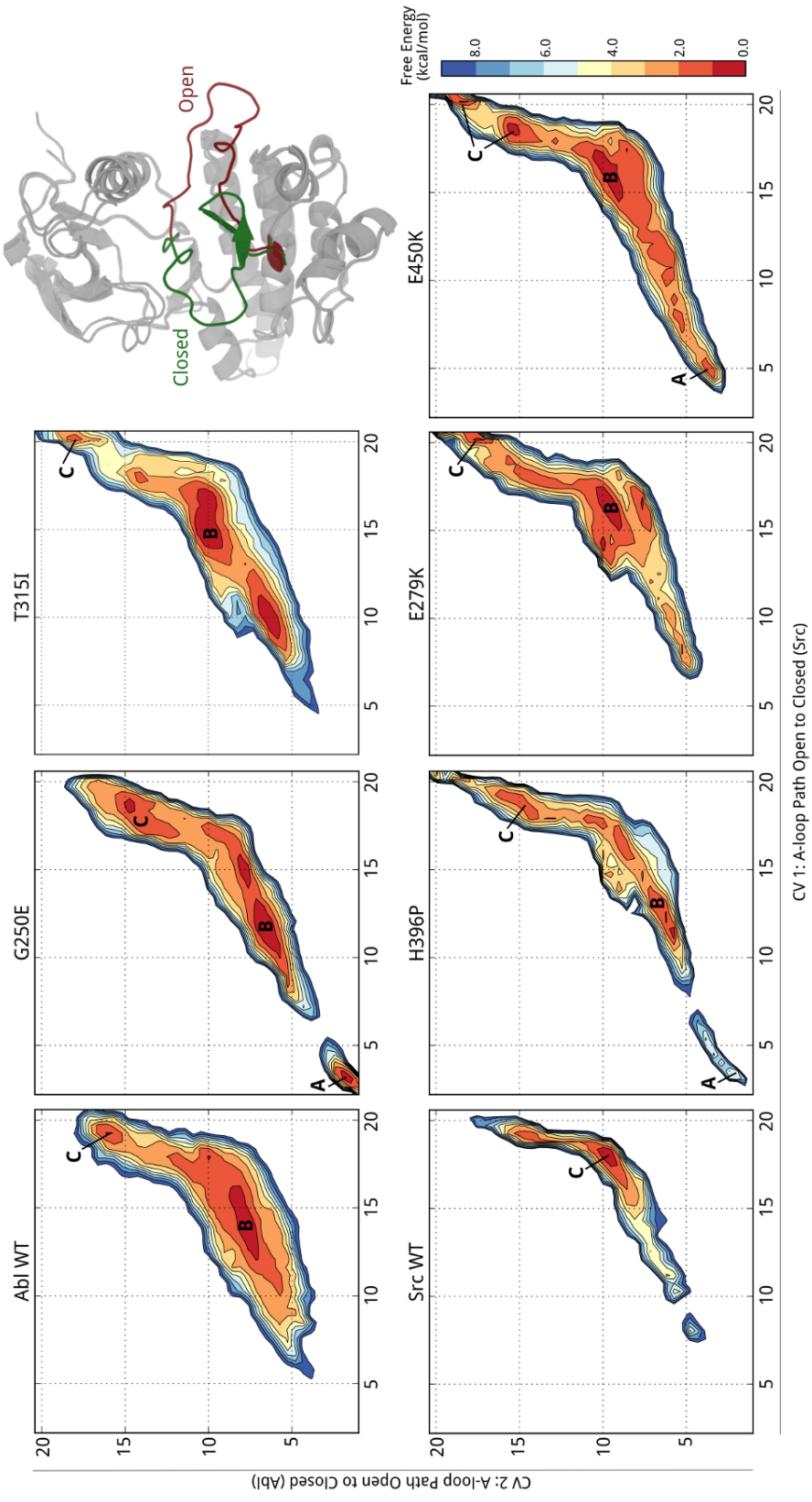

Figure 6.11: Reweighted FESs of Abl WT, Src and the resistant mutants of Abl. Projections along two path CVs describing the opening of the A-loop in Src (CV1) and Abl (CV2). The energy minimum corresponding to open, semi-open and closed A-loop conformations, when present, are labelled with A, B and C, respectively. The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. Figure taken from Ref [93].

Nonetheless, a striking difference can be observed in the FESs of the resistant mutants. If for Abl WT and Src the deepest minima correspond to the semi-open or closed conformation of the A-loop (see Figure 6.11, B or C respectively), all mutants show a major variability and a more complex landscape, with multiple metastable minima and A-loop "open like" conformations partially stabilised. Furthermore, in G250E, E450K and to a lesser extend in H396P, E279K and T315I, a minimum appears in correspondence of the fully open A-loop conformation (basin A, Figure 6.11), proving that as observed for oncogenic mutants[100], more open A-loop conformations are also explored by the imatinib resistant mutants, even in the non phosphorylated form.

For E450K fully open A-loop conformations are visited with a barrier of 3 kcal/mol, while for H396P completely open conformations (CV1 values lower than 5, see Figure 6.11) are sampled but not fully stable. E279K and T315I seem to present an opening tendency as well, with a landscape more similar to the one of Src with high barriers to cross. The T315I mutant, like Abl WT, has a single broad minimum, that corresponds to the semi-open A-loop conformation (see Figure 6.11). It seems that just G250E, E450K and H396P are able to overcome the high barriers towards a fully open A-loop conformation. This could be probably correlated with the unfolding events of the αG and αC helices, previously described. Indeed, referring to literature, conformational transitions could be associated with the unfolding of crucial structural elements ("cracking" points[216, 101], see Introduction 1.5.1).

6.7 Binding free-energy of imatinib to Abl and T315I mutant

Among all the resistant mutants considered in the study, the "gatekeeper" T315I is a remarkable exception. Not only because it is known to affect directly the binding of imatinib, but also because it does not alter the free energy of the DFG flip and it shows just a weak tendency to visit open A-loop conformations. However, it extensively samples the DFG-in active conformation, as shown in Figure 6.4 by the broad minimum of the FES. To better understand the mode of action of the T315I mutant and to verify the recently proposed induced fit theory[80, 79], the (un-)binding free energy profile of imatinib to both Abl WT and T315I was reconstructed (see Section 2.1.6).

Figure 6.12 illustrates the important differences between the binding mechanism in the two kinases. In both cases, the deepest minimum of the FES (basin A) corresponds to the kinase bound to imatinib, in the conformation seen in the crystal structure. In Abl WT only one unbinding pathway can be identified, characterised by an additional metastable minimum which corresponds to an external binding pose, where imatinib slides below the αC-helix (basin B, Figure 6.12). It is noticeable that when imatinib approaches the binding site (basin C), the DFG motif is already in the DFG-out conformation thus, validating the theory of a conformational selection mechanism[20, 78, 187].

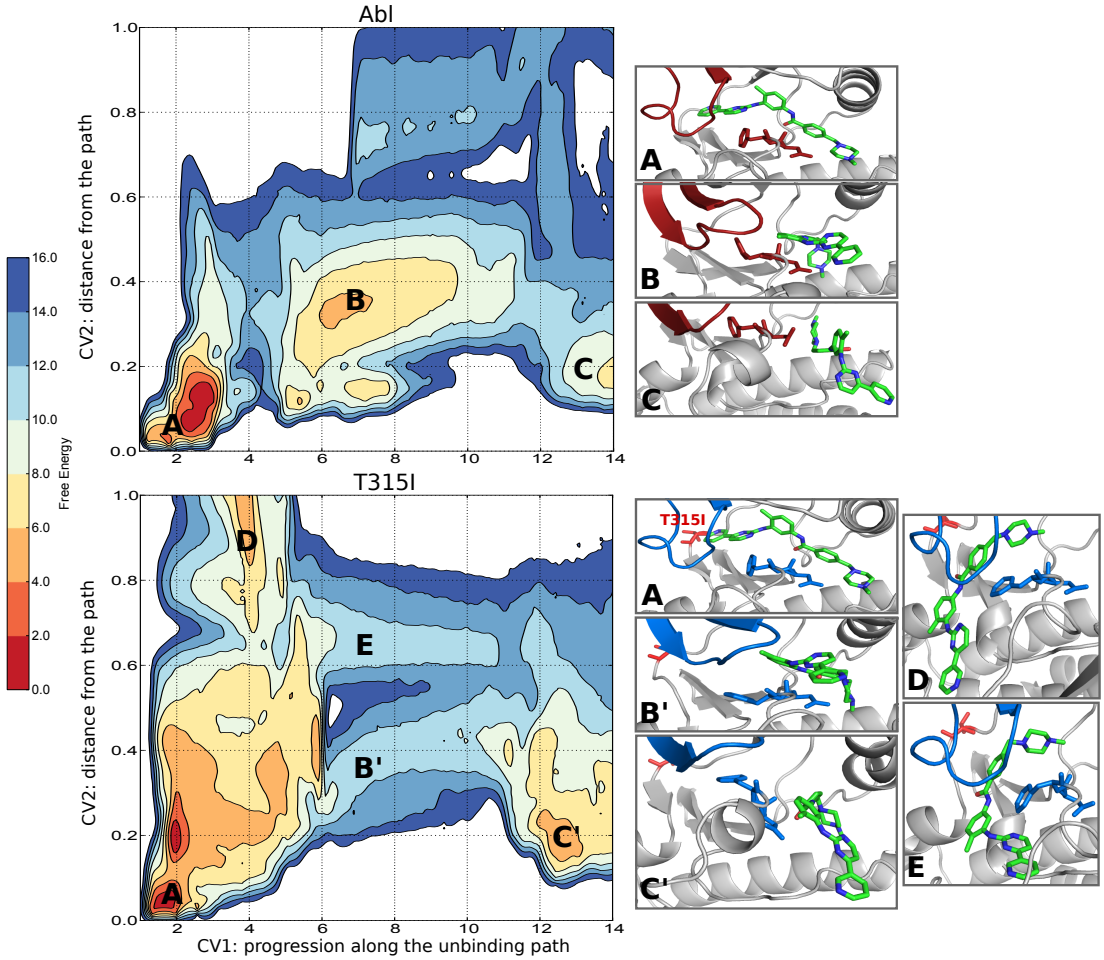

Figure 6.12: Reconstruction of the (un-)binding FES of imatinib to Abl and T315I. The structures corresponding to the local minima and transition states of the FES are identified with letters. Shown in green stick is the drug imatinib. In the structures of the T315I, the mutation is shown in red sticks. The error on the minima and barriers of the FESs amounts to 0.5 kcal/mol. Figure taken from Ref [93].

On the contrary, in the gatekeeper mutant not only is the barrier to unbinding 2 kcal/mol lower than in the WT, but imatinib has two distinct unbinding pathways (via

B' or via E, Figure 6.12). The first pathway resembles the one of Abl WT (via B', Figure 6.12), however the shift from the external pose (basin C') to the docked one is hampered by a higher energy barrier with no stable intermediate states (B'). The energy gain for T315I in the passage from C' to A' is lower than for Abl WT (ΔG of 6 and 8, respectively) thus, disfavouring the binding and in agreement with the higher imatinib IC_{50} value registered for the mutant. The alternative pathway for the gatekeeper mutant (via D and E), sees imatinib sliding below the P-loop and binding to the kinase through the ATP site. This pathway was originally proposed as the main mechanism of binding of type I inhibitors while, for type II inhibitors both binding pathways were thought to be feasible[217].

As mentioned previously, an induced fit mechanism has been proposed for Abl WT, involving the P-loop, and responsible for the high binding affinity towards imatinib[80, 79]. The simulations instead point towards a two step mechanism of binding of imatinib to Abl WT: a conformational selection step followed by an induced fit. When imatinib approaches the binding site (basin C, Figure 6.12), the kinase already adopts the DFG-out conformation (conformational selection) and subsequently, the main structural change between basin B and the more stable basin A is effectively the conformational change of the P-loop, with the adoption of the kinked conformation (induced fit). In the "gatekeeper" mutant, both mechanisms seem to be affected by the mutation. On the one hand, the DFG-out conformation is destabilised by the adoption of the Asp-down intermediate (basin C', Figure 6.12). On the other hand, the more variable landscape, close to basin A, is characterised by a broader metastable minimum, symptom of a general instability of imatinib in the binding site.

6.8 Discussion

The analyses performed in this chapter shed light on the multiple mechanisms of several imatinib resistant mutants: the destabilisation of the inactive DFG-out conformation together with the opening of the A-loop favour the shift towards the kinase active conformation and lead to a direct ATP competition. Eventually, every mutant would adopt one or both mechanisms to lower imatinib affinity towards Abl WT.

The RMSF analyses show a significant trend, for which the mutations conferring resistance generally determine a reduction of the sub-μs dynamics. The equilibrium

between the DFG-in and DFG-out conformations has been proved to be a delicate balance between entropic and enthalpic contributions, as seen for the G250E and T315I mutants. In this perspective, the changes in the sub-μs dynamics could be predictive of the DFG-out conformation energetics.

As for Src, the study of the mutants of Abl was helpful for the understanding of the allosteric connections regulating the Abl KD (see Figure 6.13). In agreement with the allosteric coupling analysis (see Section 4.2), the A-loop and the αC-helix of Abl result to be strongly correlated, proved by the cross-effect of the E279K and H296P mutations. The P-loop, whose mutation G250E affects the A-loop dynamics and the N-lobe loops in general as well takes part in this interplay. The αG-helix, on the contrary, seems to play a marginal role in Abl, as mutants in it do not rigidify the A-loop, and have the weakest effect on imatinib IC_{50}.

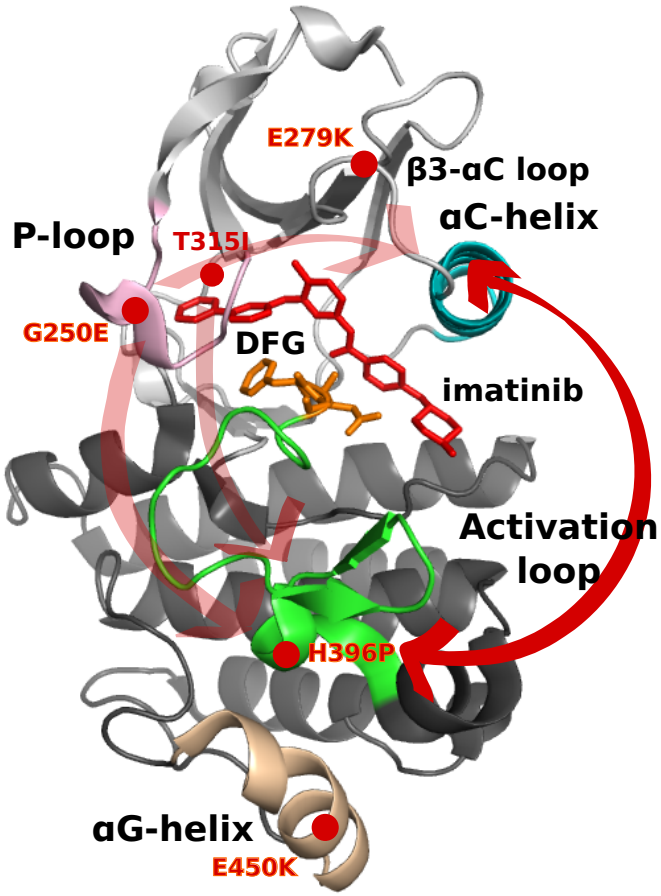

Figure 6.13: Structure with highlighted by red dots the location of the resistant mutants of Abl. Red arrows illustrate the dominant allosteric pathways identified by the analyses performed. The connection between the αC-helix and the A-loop is a feature of the Abl structure.

Finally, the (un-)binding free energies of imatinib to Abl WT and to T315I have unveiled both mechanisms of binding. Indeed, contrary to what was stated in recent papers[79, 80], imatinib binds to Abl WT first by selecting the DFG-out conformation and subsequently the binding is stabilised by an induced fit, with the P-loop adopting the kinked conformation. The loss of both mechanisms in the "gatekeeper" mutant, together with the loss of an hydrogen bond crucial for the stabilisation of the binding[91, 61], are in agreement with the lower affinity of imatinib for the mutant.

Chapter 7

Conclusions

In this thesis, I have investigated the molecular mechanisms responsible for the observed differences in imatinib activity towards homologous tyrosine kinases and the reasons leading to drug resistance. By using a combination of computational and experimental approaches I found a link between sub-μs dynamics and the conformational dynamics involved in the regulation of tyrosine kinases (TKs) and in the binding of the drug imatinib.

In particular, the dynamics of specific functional motifs in TKs, such as the P-loop, αC-helix and the A-loop are involved in the binding process of drugs. Not only because they are close to the drug binding site, but also because they are directly or indirectly (allosterically) involved in the conformational changes that lead to the adoption of the druggable conformations (DFG-in for type I inhibitors, DFG-out for type II, opening of the A-loop, etc.).

The sub-μs dynamics was found to be correlated to the conformational dynamics of this group of kinases. Therefore, the former is able to discriminate binders from non binders among highly homologous TKs. Indeed, the structural plasticity is able to influence the large scale domain motions, and influences the free energy penalty of the DFG-out conformation. Among the kinases studied, Abl (the main target of imatinib) is very flexible, being able to explore a wider range of conformations, mainly due to the specific rearrangements of the αC-helix, not observed in other TKs and lost in the drug-resistant mutants of Abl. Moreover, the dynamics of the functionally-relevant motifs are highly correlated in Abl, and in particular they are all linked to the flexible

A-loop. All these elements together concur to stabilise the DFG-out conformation, thus facilitating the binding of imatinib.

The fact that most imatinib resistant mutants of Abl affect flexible regions confirms the role of flexibility. The engineered mutants of Src confirm the link between flexibility and imatinib affinity. However, the change in the binding affinity of imatinib to the engineered mutants was smaller than expected. This might be partially due to the fact that none of the mutants are significantly affected by the A-loop flexibility. Indeed, the observed changes in the sub-μs dynamics in the clinically relevant mutants and in some of the TKs are likely to be a symptom of a more profound change of the conformational free energy landscape. Two mechanisms have been identified: the destabilisation of the inactive DFG-out conformation and the opening of the A-loop, with a shift towards the kinase active conformation.

The changes in the sub-μs dynamics reflect (and in part determine) the global changes of the conformational landscape. More flexible TKs favour the DFG flip and a larger DFG-out population, while rigid TKs favour the DFG-in active conformation. The phosphorylation of the A-loop reduces the activity of type II drugs (such as imatinib) but not of type I drugs, as mentioned in Section 1.4.2 of the Introduction, because it induces a shift towards DFG-in (type I target), lowering the population of DFG-out (type II target). Moreover, the effect of the phosphorylation in lowering imatinib affinity is much more evident in flexible kinases (as Abl and Kit) than in the more rigid ones (as Src). This could be explained by the fact that flexible kinases populate the DFG-out conformations rather than the rigid TKs, so that the population shift for the latter will be minimum and therefore not changing the already low imatinib activity.

The in-out equilibrium is a delicate balance between entropic and enthalpic contributions, easily affected by even slight changes in the dynamics, as proved in the resistant mutants of Abl. Overall, our analysis points to the fact that imatinib would select for the DFG-out conformation, but its adoption and stability is influenced by a complex interplay among allostery, sub-μs dynamics and large conformational changes.

Finally, this work has also helped clarifying the mechanism of binding of imatinib to the Abl T315I "gatekeeper" mutant. In the case of Abl, a two-step mechanism of binding has been outlined. Imatinib will first select for the DFG-out conformation and, subsequently, the adoption of the kinked conformation of the P-loop will con-

fer high stability to the binding. This two-step mechanism has been impaired in the "gatekeeper" mutant T315I, resulting in a destabilisation of the binding.

On the whole, given the links between sub-μs dynamics and the conformational changes affecting the inhibitors activity, the former sub-μs dynamics might be used to predict the effect of mutations on drug binding and allow for the rational drug design of novel TKs inhibitors more resilient to the emergence of resistance.

Bibliography

[1] Rina Barouch-Bentov and Karsten Sauer. Mechanisms of drug resistance in kinases. *Expert Opin. Investig. Drugs*, 20(2):153–208, February 2011.

[2] Hains Frauenfelder, Stephen G. Sligar, and Peter G. Wolynes. The Energy Landscapes and Motions of Proteins. *Science*, 254:1598–1603, 1991.

[3] Katherine A. Henzler-Wildman, Vu Thai, Ming Lei, Maria Ott, Magnus Wolf-Watz, Tim Fenn, Ed Pozharski, Mark A. Wilson, Gregory A. Petsko, Martin Karplus, Christian G. Hübner, and Dorothee Kern. Intrinsic motions along an enzymatic reaction trajectory. *Nature*, 450:838–44, 2007.

[4] Horst Kunz. Emil Fischer–unequalled classicist, master of organic chemistry research, and inspired trailblazer of biological chemistry. *Angewandte Chemie (International ed. in English)*, 41(23):4439–51, December 2002.

[5] Danile E Koshland Jr. Properties of the Active Site of Enzymes. *Annuals New York Academy of Science*, pages 630–642, 1963.

[6] J. A. Hartsuck, M. L. Ludwig, H. Muirhead, T. A. Steitz, and W. N. Lipscomb. Carboxypeptidase A, II. The three-dimensional electron density map at 6 Å resolution. *PNAS*, 53:396–403, 1964.

[7] T.A. Steitz, M. Shoham, and W.S. Bennett Jr. Structural dynamics of yeast hexokinase during catalysis. *Phil. Trans. R. Soc. Lond.*, 293:43–52, 1981.

[8] Mark Gerstein, Arthur M. Lesk, and Cyrus Chothia. Structural Mechanisms for Domain Movements in Proteins. *Biochemistry*, 33(22):6739–6749, June 1994.

[9] Robert G. Smock and Lila M. Gierasch. Sending Signals Dynamically. *Science*, 324:198–203, 2009.

[10] Joanna F. Swain and Lila M. Gierasch. The changing landscape of protein allostery. *Current opinion in structural biology*, 16(1):102–8, February 2006.

[11] Elan Z. Eisenmesser, Oscar Millet, Wladimir Labeikovsky, Dmitry M. Korzhnev, Magnus Wolf-Watz, Daryl A. Bosco, Jack J. Skalicky, Lewis E. Kay, and Dorothee Kern. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 438(7064):117–21, November 2005.

[12] Kurt Wüthrich. NMR studies of structure and function of biological macromolecules. *Journal of Biomolecular NMR*, 27(1):235–267, September 2002.

[13] Katherine A. Henzler-Wildman, Ming Lei, Vu Thai, S. Jordan Kerns, Martin Karplus, and Dorothee Kern. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 450:913–916, November 2007.

[14] Navratna Vajpai, André Strauss, Gabriele Fendrich, Sandra W Cowan-Jacob, Paul W Manley, Stephan Grzesiek, and Wolfgang Jahnke. Solution Conformations and Dynamics of ABL Kinase-Inhibitor Complexes Determined by NMR Substantiate the Different Binding Modes of Imatinib/Nilotinib and Dasatinib. *Journal of Biological Chemistry*, 283(26):18292–18302, 2008.

[15] Lukasz Skora, Jürgen Mestan, Doriano Fabbro, Wolfgang Jahnke, and Stephan Grzesiek. NMR reveals the allosteric opening and closing of Abelson tyrosine kinase by ATP-site and myristoyl pocket inhibitors. *PNAS*, 110(47):E4437–45, November 2013.

[16] R. M. Daniel, R. V. Dunn, J. L. Finney, and J. C. Smith. The role of dynamics in enzyme activity. *Annual Review of Biophysics and Biomolecular Structure*, 32:69–92, January 2003.

[17] Zachariah H. Foda, Yibing Shan, Eric T. Kim, David E. Shaw, and Markus A. Seeliger. A dynamically coupled allosteric network underlies binding cooperativity in Src kinase. *Nat. Comm.*, 6:1–10, 2015.

[18] Jacques Monod, Jean-Pierre Changeux, and François Jacob. Allosteric proteins and cellular control systems. *Journal of Molecular Biology*, 6:306–329, April 1963.

[19] J Monod, J Wyman, and J P Changeux. On the Nature of Allosteric Transitions: a Plausible Model. *Journal of Molecular Biology*, 12:88–118, May 1965.

[20] Markus A Seeliger, Bhushan Nagar, Filipp Frank, Xiaoxian Cao, M Nidanie Henderson, and John Kuriyan. c-Src binds to the cancer drug imatinib with an inactive Abl/c-Kit conformation and a distributed thermodynamic penalty. *Structure*, 15(3):299–311, March 2007.

[21] Zhengshuang Shi, Katheryn A. Resing, and Natalie G. Ahn. Networks for the allosteric control of protein kinases. *Current opinion in structural biology*, 16(6):686–92, December 2006.

[22] Alexandr P Kornev and Susan S Taylor. Defining the conserved internal architecture of a protein kinase. *Biochimica et Biophysica Acta*, 1804(3):440–4, March 2010.

[23] Alexandr P Kornev, Susan S Taylor, and Lynn F Ten Eyck. A helix scaffold for the assembly of active protein kinases. *PNAS*, 105(38):14377–82, September 2008.

[24] Alexandr P Kornev, Nina M Haste, Susan S Taylor, and Lynn F Ten Eyck. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *PNAS*, 103(47):17783–17788, 2006.

[25] Alexey Aleshin and Richard S Finn. SRC: A Century of Science Brought to the Clinic. *Neoplasia*, 12(8):599–607, 2010.

[26] Jeffrey A Ubersax and James E Ferrell. Mechanisms of specificity in protein phosphorylation. *Nature reviews. Molecular cell biology*, 8(7):530–41, July 2007.

[27] Martinez R. Manning G., Whyte D.B. The protein kinase complement of the human genome. *Science*, 298:1912–1934, 2002.

[28] Martin E. M. Noble, Jane A. Endicott, and Louise N. Johnson. Protein kinase inhibitors: insights into drug design from structure. *Science*, 303:1800–1804, 2004.

[29] D. J. Matthews and M. E. Gerritsen. *Targeting protein kinases for cancer therapy*. Wiley, 2011.

[30] Rosalind H. Gunby, Elisa Sala, Carmen J. Tartari, Miriam Puttini, Carlo Gambacorti-Passerini, and Luca Mologni. Oncogenic Fusion Tyrosine Kinases as Molecular Targets for Anti-Cancer Therapy. *Anti-Cancer Agents in Medicinal Chemistry*, 7:594–611, 2007.

[31] Douglas Hanahan and Robert A. Weinberg. The Hallmarks of Cancer. *Cell*, 100:57–70, 2000.

[32] Douglas Hanahan and Robert A. Weinberg. Hallmarks of Cancer : The Next Generation. *Cell*, 144(5):646–674, 2011.

[33] Peter K. Vogt. Retroviral Oncogenes: A Historical Primer. *Nat Rev Cancer*, 12(9):639–648, 2012.

[34] M. A. Young, S. Gonfloni, G. Superti-Furga, B. Roux, and J. Kuriyan. Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell*, 105(1):115–26, April 2001.

[35] John C. Williams, Albert Weijland, Stefania Gonfloni, Andy Thompson, Sara A. Courtneidge, Giulio Superti-Furga, and Rik K. Wierenga. The 2.35 Å Crystal Structure of the Inactivated Form of the Inactivated form of Chicken Src: A Dynamic Molecule with Multiple Regulatory Interactions. *J Mol Biol*, 274:757–775, 1997.

[36] Wenqing Xu, Stephen C. Harrison, and Micheal J. Eck. Three-dimensional structure of the tyrosine kinase c-Src. *Nature*, 385:595–602, 1997.

[37] W. Xu, A. Doshi, M. Lei, M. J. Eck, and S. C. Harrison. Crystal structures of c-Src reveal features of its autoinhibitory mechanism. *Molecular Cell*, 3(5):629–38, May 1999.

[38] Morgan Huse and John Kuriyan. The Conformational Plasticity of Protein Kinases. *Cell*, 109:275–282, 2002.

[39] A. G. Tatosyan and O. A. Mizenina. Kinases of the Src family: structure and functions. *Biochemistry*, 65(1):49–58, January 2000.

[40] Michael W. N. Deininger, John M. Goldman, and Junia V. Melo. The molecular biology of chronic myeloid leukemia. *Blood*, 96(10):3343–3356, 2000.

[41] Nicole Dölker, Maria W Górna, Ludovico Sutto, Antonio S Torralba, Giulio Superti-Furga, and Francesco L Gervasio. The SH2 Domain Regulates c-Abl Kinase Activation by a Cyclin-Like Mechanism and Remodulation of the Hinge Motion. *PLoS Comput. Biol.*, 10(10):1–12, October 2014.

[42] Bhushan Nagar, Oliver Hantschel, Markus Seeliger, Jason M Davies, William I Weis, Giulio Superti-Furga, and John Kuriyan. Organization of the SH3-SH2 unit in active and inactive forms of the c-Abl tyrosine kinase. *Molecular Cell*, 21(6):787–98, March 2006.

[43] Bhushan Nagar, Oliver Hantschel, Matthew A Young, Klaus Scheffzek, Darren Veach, William Bornmann, Bayard Clarkson, Giulio Superti-Furga, and John Kuriyan. Structural Basis for the Autoinhibition of c-Abl Tyrosine Kinase. *Cell*, 112:859–871, 2003.

[44] Charlotte E. Edling and Bengt Hallberg. c-Kit - a hematopoietic cell essential receptor tyrosine kinase. *The International Journal of Biochemistry & Cell Biology*, 39(11):1995–8, January 2007.

[45] Alain B. Schreiber, Towia A. Libermann, Irit Lax, Yosef Yarden, and Joseph Schlessinger. Biological Role of Epidermal Growth Factor-Receptor Clustering. *The Journal of Biological Chemistry*, 258(2):846–853, 1983.

[46] Alessandra Gentile, Livio Trusolino, and Paolo M Comoglio. The Met tyrosine kinase receptor in development and cancer. *Cancer Metastasis Review*, 27(1):85–94, March 2008.

[47] U. Lorenz, K. S. Ravichandran, D. Pei, C. T. Walsh, S. J. Burakoff, and B. G. Neel. Lck-dependent tyrosyl phosphorylation of the phosphotyrosine phosphatase SH-PTP1 in murine T cells. *Molecular and Cellular Biology*, 14(3):1824–34, March 1994.

[48] Carmen Birchmeier, Walter Birchmeier, Ermanno Gherardi, and George F Vande Woude. Met, metastasis, motility and more. *Nature Reviews. Molecular Cell Biology*, 4(12):915–25, December 2003.

[49] Y. Yarden and M. X. Sliwkowski. Untangling the ErbB signalling network. *Nature reviews. Molecular Cell Biology*, 2(2):127–37, February 2001.

[50] Nancy E. Hynes and Heidi A. Lane. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nature Reviews. Cancer*, 5(5):341–54, May 2005.

[51] Doriano Fabbro, Stephan Ruetz, Elisabeth Buchdunger, Sandra W Cowan-Jacob, Gabriele Fendrich, Janis Liebetanz, Jürgen Mestan, Terence O'Reilly, Peter Traxler, Bhabatosh Chaudhuri, Heinz Fretz, Jürg Zimmermann, Thomas Meyer, Giorgio Caravatti, Pascal Furet, and Paul W Manley. Protein kinases as targets for anticancer agents: from inhibitors to useful drugs. *Pharmacology & Therapeutics*, 93:79–98, February 2002.

[52] Philip Cohen. Protein kinases - the major drug targets of the twenty-first century? *Nature Review*, 1:309–315, 2002.

[53] Renaud Capdeville, Elisabeth Buchdunger, Juerg Zimmermann, and Alex Matter. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat. Rev. Drug Discov.*, 1(7):493–502, July 2002.

[54] G. Daley, R. Van Etten, and David Baltimore. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science*, 247:824–830, February 1990.

[55] Oliver Hantschel and Giulio Superti-Furga. Regulation of the c-Abl and Bcr-Abl tyrosine kinases. *Nat. Rev. Mol. Cell Biol.*, 5(1):33–44, January 2004.

[56] E. Buchdunger, A. Matter, and B. J. Druker. Bcr-Abl inhibition as a modality of CML therapeutics. *Biochimica et biophysica acta*, 1551(1):M11–8, 2001.

[57] Hagop M Kantarjian, Jorge E Cortes, Susan O Brien, Francis Giles, Guillermo Garcia-Manero, Stefan Faderl, Deborah Thomas, Sima Jeha, Mary Beth Rios, Laurie Letvak, Kathy Bochinski, Ralph Arlinghaus, and Moshe Talpaz. Imatinib

mesylate therapy in newly diagnosed patients with Philadelphia chromosome - positive chronic myelogenous leukemia: high incidence of early complete and major cytogenetic responses. *Blood*, 101:97–100, 2003.

[58] Michele Baccarani, Giuseppe Saglio, John Goldman, Andreas Hochhaus, Bengt Simonsson, Frederick Appelbaum, Jane Apperley, Francisco Cervantes, Jorge Cortes, Michael Deininger, François Guilhot, Mary Horowitz, Timothy Hughes, Hagop Kantarjian, Richard Larson, Dietger Niederwieser, Richard Silver, Rudiger Hehlmann, Washington Dc, and Alois Gratwohl. Evolving concepts in the management of chronic myeloid leukemia : recommendations from an expert panel on behalf of the European LeukemiaNet. *Blood*, pages 1809–1820, 2006.

[59] Hagop Kantarjian, Susan O'Brien, Guillermo Garcia-Manero, Stefan Faderl, Farhad Ravandi, Elias Jabbour, Jianqin Shan, and Jorge Cortes. Very long-term follow-up results of imatinib mesylate therapy in chronic phase chronic myeloid leukemia after failure of interferon alpha therapy. *Cancer*, 118(12):3116–22, June 2012.

[60] Michael W N Deininger and Brian J Druker. Specific Targeted Therapy of Chronic Myelogenous Leukemia with Imatinib. *Pharmacol Rev*, 55(3):401–423, 2003.

[61] Shun J. Lee and Jean Y. J. Wang. Exploiting the promiscuity of imatinib. *Journal of Biology*, 8(30):1–4, January 2009.

[62] T. Schindler, W. Bornmann, P. Pellicena, W. T. Miller, B. Clarkson, and J. Kuriyan. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science*, 289:1938–42, 2000.

[63] E. Buchdunger, C. L. Cioffi, N. Law, D. Stover, S. Ohno-Jones, B. J. Druker, and N. B. Lydon. Abl protein-tyrosine kinase inhibitor STI571 inhibits in vitro signal transduction mediated by c-kit and platelet-derived growth factor receptors. *The Journal of pharmacology and experimental therapeutics*, 295(1):139–45, 2000.

[64] Yasmine Asses, Vincent Leroux, Safia Tairi-Kellou, Rosanna Dono, Flavio Maina, and Bernard Maigret. Analysis of c-Met Kinase Domain Complexes:

A New Specific Catalytic Site Receptor Model for Defining Binding Modes of ATP-Competitive Ligands. *Chemical Biology And Drug Design*, 74:560–570, 2009.

[65] Michael Deininger, Elisabeth Buchdunger, and Brian J. Druker. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood*, 105(7):2640–53, April 2005.

[66] Jürg Zimmermann, Elisabeth Buchdunger, Helmut Mett, Thomas Meyer, and Nicholas B. Lydon. Potent and selective inhibitors of the Abl-kinase: phenylamino-pyrimidine (PAP) derivatives. *Bioorganic & Medicinal Chemistry Letters*, 7(2):187–192, 1997.

[67] Bhushan Nagar, William G Bornmann, Patricia Pellicena, Thomas Schindler, Darren R Veach, W Todd Miller, Bayard Clarkson, and John Kuriyan. Crystal Structures of the Kinase Domain of c-Abl in Complex with the Small Molecule Inhibitors PD173955 and Imatinib. *Cancer Research*, 62:4236–4243, 2002.

[68] Clifford D. Mol, Douglas R. Dougan, Thomas R. Schneider, Robert J. Skene, Michelle L. Kraus, Daniel N. Scheibe, Gyorgy P. Snell, Hua Zou, Bi-Ching Sang, and Keith P. Wilson. Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *The Journal of Biological Chemistry*, 279(30):31655–63, July 2004.

[69] Marc D. Jacobs, Paul R. Caron, and Brian J. Hare. Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: Structure of lck/imatinib complex. *Proteins*, pages 1451–60, 2007.

[70] Yibing Shan, Markus A. Seeliger, Michael P Eastwood, Filipp Frank, Huafeng Xu, Morten Ø. Jensen, Ron O. Dror, John Kuriyan, and David E Shaw. A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc. Natl. Acad. Sci. USA*, 106(1):139–144, 2009.

[71] Alexey Aleksandrov and Thomas Simonson. A Molecular Mechanics Model for Imatinib and Imatinib:Kinase Binding. *Journal of Computational Chemistry*, 31:1550–1560, 2009.

[72] https://commons.wikimedia.org/wiki/File:Imatinib_in_its_binding_site.svg. [Source of Figure 1.13 shown in the Introduction].

[73] Yilin Meng and Benoît Roux. Locking the Active Conformation of c-Src Kinase through the Phosphorylation of the Activation Loop. *J. Mol. Biol.*, 426:423–435, January 2014.

[74] G. M. Torrie and J. P. Valleau. Nonphysical sampling distribution in Monte Carlo free-energy estimation: Umbrella Sampling. *Journal of Computational Physics*, 23:187–199, 1977.

[75] Sanjay B. Hari, B. Gayani K. Perera, Pratistha Ranjitkar, Markus A. Seeliger, and Dustin J. Maly. Conformation-selective inhibitors reveal differences in the activation and phosphate-binding loops of the tyrosine kinases Abl and Src. *ACS Chemical Biology*, 8(12):2734–43, December 2013.

[76] Lisa M. Wodicka, Pietro Ciceri, Mindy I. Davis, Jeremy P. Hunt, Mark Floyd, Sara Salerno, Xuequn H. Hua, Julia M Ford, Robert C. Armstrong, Patrick P. Zarrinkar, and Daniel K. Treiber. Activation State-Dependent Binding of Small Molecule Kinase Inhibitors : Structural Insights from Biochemistry. *Chem. & Biol.*, 17(11):1241–1249, November 2010.

[77] Ketan S. Gajiwala, Joe C. Wu, James Christensen, Gayatri D. Deshmukh, Wade Diehl, Jonathan P. Dinitto, Jessie M. English, Michael J. Greig, You-ai He, Suzanne L. Jacques, Elizabeth A. Lunney, Michele McTigue, David Molina, Terri Quenzer, Peter A. Wells, Xiu Yu, Yan Zhang, Aihua Zou, Mark R. Emmett, Alan G. Marshall, Hui-Min Zhang, and George D. Demetri. KIT kinase mutants show unique mechanisms of drug resistance to imatinib and sunitinib in gastrointestinal stromal tumor patients. *PNAS*, 106:1542–1547, 2009.

[78] Alexey Aleksandrov and Thomas Simonson. Molecular Dynamics Simulations Show That Conformational Selection Governs the Binding Preferences of Imatinib for Several Tyrosine Kinases. *Journal of Biological Chemistry*, 285(18):13807–13815, 2010.

[79] Roman V. Agafonov, Christopher Wilson, Renee Otten, Vanessa Buosi, and Dorothee Kern. Energetic dissection of Gleevec's selectivity toward human tyrosine kinases. *Nature structural & molecular biology*, 21(10):848–53, October 2014.

[80] C. Wilson, R. V. Agafonov, M. Hoemberger, S. Kutter, A. Zorba, J. Halpin, V. Buosi, R. Otten, D. Waterman, D. L. Theobald, and D. Kern. Kinase dynamics. Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science*, 347:882–886, February 2015.

[81] Y.-L. Lin, Y. Meng, W. Jiang, and B. Roux. Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. *Proceedings of the National Academy of Sciences*, 110(5):1664–1669, January 2013.

[82] Marco D'Abramo, Obdulia Rabal, Julen Oyarzabal, and Francesco Luigi Gervasio. Conformational Selection versus Induced Fit in Kinases: The Case of PI3K-γ. *Angewandte Chemie*, 123:1–6, January 2011.

[83] Gordon G Hammes, Yu-Chu Chang, and Terrence G Oas. Conformational selection or induced fit: a flux description of reaction mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 106(33):13737–41, August 2009.

[84] Ellen Weisberg, Paul W. Manley, Sandra W. Cowan-Jacob, Andreas Hochhaus, and James D. Griffin. Second generation inhibitors of BCR- ABL for the treatment of imatinib-resistant chronic myeloid leukaemia. *Nature Review*, 7:345–356, 2007.

[85] Jianming Zhang, Francisco J. Adrián, Wolfgang Jahnke, Sandra W. Cowan-Jacob, Allen G. Li, Roxana E. Jacob, Taebo Sim, John Powers, Christine Dierks, Fangxian Sun, Gui-Rong Guo, Qiang Ding, Barun Okram, Yongmun Choi, Amy Wojciechowski, Xianming Deng, Guoxun Liu, Gabriele Fendrich, André Strauss, Navratna Vajpai, Stephan Grzesiek, Tove Tuntland, Yi Liu, Badry Bursulaya, Mohammad Azam, Paul W Manley, John R. Engen, George Q. Daley, Markus Warmuth, and Nathanael S. Gray. Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. *Nature*, 463:501–6, January 2010.

[86] C. Miething, S. Feihl, C. Mugler, R. Grundler, N. Von Bubnoff, F. Lordick, C. Peschel, and J. Duyster. The Bcr-Abl mutations T315I and Y253H do not confer a growth advantage in the absence of imatinib. *Leukemia*, 20:650–657, 2006.

[87] Carlo B. Gambacorti-Passerini, Rosalind H. Gunby, Rocco Piazza, Annamaria Galietta, and Roberta Rostagno. Molecular mechanisms of resistance to imatinib in Philadelphia-chromosome-positive leukaemias. *Lancet Oncology*, 4:75–85, 2003.

[88] Mohammad Azam, Robert R Latek, and George Q Daley. Mechanisms of Autoinhibition and STI-571 / Imatinib Resistance Revealed by Mutagenesis of BCR-ABL. *Cell*, 112:831–843, 2003.

[89] Mohammad Azam, Valentina Nardi, William C Shakespeare, Chester A Metcalf, Regine S Bohacek, Yihan Wang, Raji Sundaramoorthi, Piotr Sliz, Darren R Veach, William G Bornmann, Bayard Clarkson, David C Dalgarno, Tomi K Sawyer, and George Q Daley. Activity of dual SRC-ABL inhibitors highlights the role of BCR/ABL kinase dynamics in drug resistance. *Proc. Natl. Acad. Sci. USA*, 103(24):9244–9, June 2006.

[90] Patrick A. Eyers. Getting to grips with drug resistance in the human protein kinase superfamily. *European Pharmaceutical Review*, 18:49–54, 2013.

[91] M. E. Gorre, M. Mohammed, K. Ellwood, N. Hsu, R. Paquette, P. N. Rao, and C. L. Sawyers. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science*, 293:876–80, August 2001.

[92] Neil P Shah, John M Nicoll, Bhushan Nagar, Mercedes E Gorre, Ronald L Paquette, John Kuriyan, and Charles L Sawyers. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell*, 2:117–125, 2002.

[93] Silvia Lovera, Maria Morando, Encarna Pucheta-Martinez, Jorge L. Martinez-Torrecuadrada, Giorgio Saladino, and Francesco L. Gervasio. Towards a Molec-

ular Understanding of the Link Between Imatinib Resistance and Kinase Conformational Dynamics. *Plos Computational Biology*, 11 (11):e1004578, 2015.

[94] Susan Branford, Zbigniew Rudzki, Sonya Walsh, Ian Parkinson, Andrew Grigg, Jeff Szer, Kerry Taylor, Richard Herrmann, John F Seymour, Chris Arthur, David Joske, Kevin Lynch, and Tim Hughes. Detection of BCR-ABL mutations in patients with CML treated with imatinib is virtually always accompanied by clinical resistance, and mutations in the ATP phosphate-binding loop (P-loop) are associated with a poor prognosis. *Blood*, 102(1):276–83, July 2003.

[95] F-X Mahon. An amino-acid switch in the BCR-ABL kinase domain modifies sensitivity to imatinib mesylate. *Leukemia*, 19:1671–73, 2005.

[96] Andrea Hoelbl, Christian Schuster, Boris Kovacic, Bingmei Zhu, Mark Wickre, Maria A Hoelzl, Sabine Fajmann, Florian Grebien, Wolfgang Warsch, Gabriele Stengl, Lothar Hennighausen, Valeria Poli, Hartmut Beug, Richard Moriggl, and Veronika Sexl. Stat5 is indispensable for the maintenance of bcr/abl-positive leukaemia. *EMBO Molecular Medicine*, 2(3):98–110, March 2010.

[97] Frank D. Böhmer, Luchezar Karagyozov, Andrea Uecker, Hubert Serve, Alexander Botzki, Siavosh Mahboobi, and Stefan Dove. A single amino acid exchange inverts susceptibility of related receptor tyrosine kinases for the ATP site inhibitor STI-571. *The Journal of Biological Chemistry*, 278(7):5148–55, March 2003.

[98] Anshuman Dixit and Gennady M. Verkhivker. Hierarchical modeling of activation mechanisms in the ABL and EGFR kinase domains: thermodynamic and mechanistic catalysts of kinase activation by cancer mutations. *PLoS Comput. Biol.*, 5(8):1–22, August 2009.

[99] Anshuman Dixit and Gennady M. Verkhivker. Computational modeling of allosteric communication reveals organizing principles of mutation-induced signaling in ABL and EGFR kinases. *PLoS Comput. Biol.*, 7(10):1–19, October 2011.

[100] Ludovico Sutto and Francesco Luigi Gervasio. Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proc. Natl. Acad. Sci. USA*, 110(26):10616–21, June 2013.

[101] Yibing Shan, MichaelăP. Eastwood, Xuewu Zhang, Eric T. Kim, Anton Arkhipov, Ron O. Dror, John Jumper, John Kuriyan, and David E. Shaw. Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. *Cell*, 149(4):860–870, May 2012.

[102] Kristen A. Marino, Ludovico Sutto, and Francesco Luigi Gervasio. The Effect of a Widespread Cancer-Causing Mutation on the Inactive to Active Dynamics of the B-Raf Kinase. *J Am Chem Soc*, 137:5280–5283, April 2015.

[103] Tom D Bunney, Shunzhou Wan, Nethaji Thiyagarajan, Ludovico Sutto, Sarah V Williams, Paul Ashford, Hans Koss, Margaret A Knowles, Francesco L Gervasio, Peter V Coveney, and Matilda Katan. The Effect of Mutations on Drug Sensitivity and Kinase Activity of Fibroblast Growth Factor Receptors: A Combined Experimental and Theoretical Study. *EBioMedicine*, 2(3):194–204, March 2015.

[104] Paul Whitford, Osamu Miyashita, Yaakov Levy, and Jose N Onuchic. Conformational transitions of Adenylate Kinase: switching by cracking. *J Mol Biol*, 366(5):1661–1671, 2007.

[105] Diego U. Ferreiro, Joseph A. Hegler, Elizabeth A. Komives, and Peter G. Wolynes. Localizing frustration in native proteins and protein assemblies. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50):19819–24, December 2007.

[106] Kei-Ichi Okazaki, Nobuyasu Koga, Shoji Takada, Jose N Onuchic, and Peter G Wolynes. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 103(32):11844–9, August 2006.

[107] Ravin J. Garg, Hagop Kantarjian, Susan O'Brien, Alfonso Quintás-Cardama, Stefan Faderl, Zeev Estrov, and Jorge Cortes. The use of nilotinib or dasatinib after failure to 2 prior tyrosine kinase inhibitors: long-term follow-up. *Blood*, 114:4361–8, November 2009.

[108] John S. Tokarski, John A. Newitt, Chieh Ying J. Chang, Janet D. Cheng, Michael Wittekind, Susan E. Kiefer, Kevin Kish, Francis Y. F. Lee, Robert Borzillerri, Louis J. Lombardo, Dianlin Xie, Yaqun Zhang, and Herbert E. Klei. The Structure of Dasatinib (BMS-354825) Bound to Activated ABL Kinase Domain Elucidates Its Inhibitory Activity against Imatinib-Resistant ABL Mutants. *Cancer Research*, 66:5790–97, 2006.

[109] Prithviraj Bose, Haeseong Park, Jawad Al-Khafaji, and Steven Grant. Strategies to circumvent the T315I gatekeeper mutation in the Bcr-Abl tyrosine kinase. *Leukemia Research Reports*, 2(1):18–20, January 2013.

[110] Wenjun Zheng, Bernard R. Brooks, and D. Thirumalai. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proc. Natl. Acad. Sci. USA*, 103(20):7664–9, 2006.

[111] S. Kumar, B. Ma, C. J. Tsai, N. Sinha, and R. Nussinov. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci.*, 9(1):10–9, January 2000.

[112] Doriano Fabbro, David Parkinson, and Alex Matter. Protein tyrosine kinase inhibitors : new treatment modalities? *Current Opinion in Pharmacology*, 2:374–381, 2002.

[113] Yi Liu and Nathanael S Gray. Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.*, 2(7):358–64, July 2006.

[114] Henrik Möbitz and Doriano Fabbro. Conformational bias: a key concept for protein kinase inhibition. *European Pharmaceutical Review*, 17(1), 2012.

[115] Jianming Zhang, Priscilla L. Yang, and Nathanael S. Gray. Targeting cancer with small molecule kinase inhibitors. *Nature Reviews*, 9:28–39, 2009.

[116] Kurt J. Cox, Carolyn D. Shomin, and Indraneel Ghosh. Tinkering outside the kinase ATP box: allosteric (type IV) and bivalent (type V) inhibitors of protein kinases. *Future Medicinal Chemistry*, 3(1):29–43, January 2011.

[117] Michael S Cohen, Chao Zhang, Kevan M Shokat, and Jack Taunton. Structural Bioinformatics-Based Design of Selective, Irreversible Kinase Inhibitors. *Science*, 308:1318–1321, 2004.

[118] David E. Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Yibing Shan, and Willy Wriggers. Atomic-Level Characterization. *Science*, 330:341–346, 2010.

[119] Giovanni Bussi, Francesco Luigi Gervasio, Alessandro Laio, and Michele Parrinello. Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *Journal of the American Chemical Society*, 128(41):13435–41, 2006.

[120] Dan Frenkel and Berend Smit. *Understanding molecular simulations: from algorithms to applications*. San Diego, California, USA: Academic Press, 2002.

[121] William C. Swope. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76(1):637–649, 1982.

[122] Verlet Loup. Computer "Experiments" on Classical Fluis. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review*, 159(1):98–103, 1967.

[123] Loup Verlet. Computer "experiments" on classical fluids. II. Equilibrium correlation functions. *Physical Review*, 165(1):201–214, 1968.

[124] M. Karplus and J. A. McCammon. Dynamics of Proteins: Elements and Function. *Ann Rev Biochem*, 53(20):263–300, 1983.

[125] Michael Levitt. Energy refinement of hen egg-white lysozyme. *Journal of Molecular Biology*, 82(3):393–420, 1974.

[126] B R Gelin and M Karplus. Side-chain torsional potentials: effect of dipeptide, protein, and solvent environment. *Biochemistry*, 18(7):1256–68, 1979.

[127] Shneior Lifson and A. Roig. On the Theory of Helix Coil Transition in Polypeptides. *The Journal of Chemical Physics*, 34(6):1963, 1961.

[128] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct. Funct. Bioinf.*, 65:712–725, 2006.

[129] Robert B. Best. Optimized Molecular Dynamics Force Fields Applied to the Helix Coil Transition of Polypeptides. *The Journal of Physical Chemistry B*, pages 9004–15, 2009.

[130] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D Joseph-McCarthy, L Kuchnir, K Kuczera, F T Lau, C Mattos, S Michnick, T Ngo, D T Nguyen, B Prodhom, W E Reiher, B Roux, M Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–616, April 1998.

[131] Alexander D. MacKerell, Michael Feig, and Charles L. Brooks. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry*, 25(11):1400–15, August 2004.

[132] Kresten Lindorff-Larsen, Paul Maragakis, Stefano Piana, Michael P Eastwood, Ron O Dror, and David E Shaw. Systematic validation of protein force fields against experimental data. *PloS One*, 7(2):1–6, 2012.

[133] David A. Case, Thomas E. Cheatham, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M. Merz, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J. Woods. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16):1668–88, December 2005.

[134] Berk Hess, Carsten Kutzner, David Van Der Spoel, Erik Lindahl, and David van Der Spoel. Gromacs 4: Algorithms for highly efficient, load-balanced, and

scalable molecular simulation. *J. Chem. Theory Comput.*, 4(3):435–447, March 2008.

[135] M. J. Harvey, G. Giupponi, and G. De Fabritiis. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.*, 5(6):1632–1639, June 2009.

[136] Bernard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187–217, 1983.

[137] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–802, December 2005.

[138] Kevin J. Bowers, Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida*, 2006.

[139] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. A smooth particle mesh Ewald method. *J Chem Phys*, 103(1995):8577–8593, 1995.

[140] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, October 1984.

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

[142] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126:1–7, 2007.

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

[144] Cameron Abrams and Giovanni Bussi. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy*, 16(1):163–199, 2014.

[145] Andrea Cavalli, Andrea Spitaleri, Giorgio Saladino, and Francesco L Gervasio. Investigating drug-target association and dissociation mechanisms using metadynamics-based algorithms. *Accounts of chemical research*, 48(2):277–85, 2015.

[146] Pilar Cossio Tejada. *Protein Physics by Advanced Computational Techniques : Conformational Sampling and Folded State Discrimination*. PhD thesis, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy, 2011.

[147] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

[148] I. G. Tironi and W. F. van Gunsteren. A molecular dynamics simulation study of chloroform. *Molecular Physics*, 83:381–403, 1994.

[149] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12562–6, October 2002.

[150] Alessandro Laio and Francesco L. Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics*, 71(12):1–22, December 2008.

[151] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters*, 100(2):1–4, January 2008.

[152] Paolo Raiteri, Alessandro Laio, Francesco Luigi Gervasio, Cristian Micheletti, and Michele Parrinello. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *Journal of Physical Chemistry B*, 110(8):3533–3539, 2006.

[153] Stefano Piana and Alessandro Laio. A bias-exchange approach to protein folding. *J. Phys. Chem. B*, 111:4553–4559, 2007.

[154] Davide Branduardi, Francesco Luigi Gervasio, and Michele Parrinello. From A to B in free energy space. *The Journal of Chemical Physics*, 126(5):054103, 2007.

[155] B. Isralewitz, M. Gao, and K. Schulten. Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology*, 11(2):224–230, 2001.

[156] G. A. Tribello, M. Ceriotti, and M. Parrinello. A self-learning algorithm for biased molecular dynamics. *Proceedings of the National Academy of Sciences*, 107(41):17509–17514, 2010.

[157] Ulrich H. E. Hansmann. Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters*, pages 140–150, 1997.

[158] E Weinan, W Ren, and E Vanden-Eijnden. String method for the study of rare events. *Physical Review B*, 66:1–4, 2002.

[159] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *The Journal of Chemical Physics*, 126(16):1–8, 2007.

[160] Peter G. Bolhuis, David Chandler, Christoph Dellago, and Phillip L. Geissler. Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annual Review of Physical Chemistry*, 53(1):291–318, 2002.

[161] John D. Chodera and Frank Noé. Markov state models of biomolecular conformational dynamics. *Current Opinion in Structural Biology*, 25:135–144, 2014.

[162] James F. Dama, Michele Parrinello, and Gregory A. Voth. Well-Tempered Metadynamics Converges Asymptotically. *Physical Review Letters*, 112(24):1–6, 2014.

[163] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and Others. PLUMED: a portable plugin for free-energy calculations with molecular dynamics. *Comp. Phys. Comm.*, 180:1961–1972, 2009.

[164] Veronique Van Speybroeck, Kristof De Wispelaere, Jeroen Van der Mynsbrugge, Matthias Vandichel, Karen Hemelsoet, and Michel Waroquier. First principle chemical kinetics in zeolites: the methanol-to-olefin process as a case study. *Chemical Society Reviews*, 43(21):7326–57, November 2014.

[165] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, pages 141–151, 1999.

[166] Modesto Orozco. A theoretical view of protein dynamics. *Chemical Society Reviews*, 43(14):5051–66, July 2014.

[167] A. Sali and T. L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, 234(3):779–815, December 1993.

[168] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L. Klepeis, Ron O. Dror, and David E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8):1950–8, June 2010.

[169] Robert B. Best and Gerhard Hummer. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B*, 113(26):9004–15, July 2009.

[170] W. L. Jorgensen. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.*, 103(2):335–340, 1981.

[171] Toshiko Ichiye and Martin Karplus. Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Structure, Function, and Genetics*, 11(3):205–217, November 1991.

[172] Andrea Amadei, Antonius B. M. Linssen, and Herman J. C. Berendsen. Essential dynamics of proteins. *Proteins: Structure, Function, and Genetics*, 17(4):412–425, December 1993.

[173] Micheal L. Connelly. Analytical Molecular Surface Calculation. *J. Appl. Cryst.*, 16:548–558, 1983.

[174] Frank Eisenhaber, Philip Lijnzaad, Patrick Argos, and Chris Sander. The Double Cubic Lattice Method: Efficient Approaches to Numerical Integration of Surface Area and Volume and to Dot Surface Contouring of Molecular Assemblies. *Journal of Computational Chemistry*, 16(3):273–284, 1994.

[175] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[176] Hui Li, Andrew D Robertson, and Jan H. Jensen. Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, 61(4):704–21, December 2005.

[177] Delphine C. Bas, David M. Rogers, and Jan H. Jensen. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins*, 73(3):765–83, November 2008.

[178] Ilya A Balabin, Weitao Yang, and David N. Beratan. Coarse-grained modeling of allosteric regulation in protein receptors. *PNAS*, 106(34):14253–58, 2009.

[179] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal*, 80(1):505–515, 2001.

[180] Ron Elber. Simulations of allosteric transitions. *Current Opinion in Structural Biology*, 21(2):167–172, 2011.

[181] Alan W. Sousa da Silva and Wim F. Vranken. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC research notes*, 5(1):1–8, 2012.

[182] M. Bonomi and M. Parrinello. Enhanced Sampling in the Well-Tempered Ensemble. *Physical Review Letters*, 104(19):1–4, May 2010.

[183] Michael Deighan, Massimiliano Bonomi, and Jim Pfaendtner. Efficient Simulation of Explicitly Solvated Proteins in the Well-Tempered Ensemble. *Journal of Chemical Theory and Computation*, 2(1):8–11, 2012.

[184] Patrick R Burney, Nathan White, and Jim Pfaendtner. Structural effects of methionine oxidation on isolated subdomains of human fibrin D and αC regions. *PloS One*, 9(1):1–10, January 2014.

[185] M. Bonomi, A. Barducci, and M. Parrinello. Reconstructing the Equilibrium Boltzmann Distribution from Well-Tempered Metadynamics. *Journal of Computational Chemistry*, 30(11):1615–21, 2009.

[186] Pratyush Tiwary and Michele Parrinello. A time-independent free energy estimator for metadynamics. *The journal of physical chemistry. B*, 119(3):736–42, January 2015.

[187] Silvia Lovera, Ludovico Sutto, Ralitza Boubeva, Leonardo Scapozza, Nicole Dölker, and Francesco Luigi Gervasio. The different flexibility of c-Src and c-Abl kinases regulates the accessibility of a druggable inactive conformation. *J. Am. Chem. Soc.*, 134(5):2496–9, February 2012.

[188] Herman J C Berendsen. Molecular Dynamics Simulations: The Limits and Beyond. In *Computational Molecular Dynamics: Challenges, Methods, Ideas*, 1998.

[189] J. K. Myers and C. N. Pace. Hydrogen bonding stabilizes globular proteins. *Biophysical journal*, 71(4):2033–9, 1996.

[190] Guillaume Lamoureux, Edward Harder, Igor V. Vorobyov, Benoît Roux, and Alexander D. MacKerell. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chemical Physics Letters*, 418:245–249, 2006.

[191] Jing Huang, Pedro E. M. Lopes, Benoît Roux, and Alexander D. MacKerell. Recent Advances in Polarizable Force Fields for Macromolecules: Microsecond Simulations of Proteins Using the Classical Drude Oscillator Model. *The journal of physical chemistry letters*, 5(18):3144–3150, 2014.

[192] M. R. Dyson and Y. Durocher. *Expression Systems*. Scion, 2007.

[193] Andrea Piserchio, Ranajeet Ghose, and David Cowburn. Optimized bacterial expression and purification of the c-Src catalytic domain for solution NMR studies. *Journal of Biomolecular NMR*, 44(2):87–93, 2009.

[194] Ramón Campos-Olivas, Marco Marenchino, Leonardo Scapozza, and Francesco L Gervasio. Backbone assignment of the tyrosine kinase Src catalytic domain in complex with imatinib. *Biomol. NMR Assign.*, 5(2):221–4, October 2011.

[195] J. C. Janson and L. Rydén. *Introduction to Protein Purification: Principles, High Resolution Methods, and Applications - Third Edition*. John Wiley & Sons, Inc., 2011.

[196] Handbook. *Recombinant Protein Purification Handbook-Principle and Methods*. GE Healthcare.

[197] Handbook. *Purifying Challenging Proteins-Principle and Methods*. GE Healthcare.

[198] Handbook. *Strategies for Protein Purification*. GE Healthcare.

[199] `http://www.genosphere-biotech.com/Antibodies-Purification.html`. [Source of Figure 2.10 shown in Materials and Methods].

[200] P Cuatrecasas, M Wilchek, and C B Anfinsen. Selective enzyme purification by affinity chromatography. *Proceedings of the National Academy of Sciences of the United States of America*, 61(2):636–643, 1968.

[201] http://oregonstate.edu/instruct/bb450/fall14/stryer7/3/figure_03_04.jpg. [Biochemistry, Seventh Edition, 2012 W.H. Freeman and Company. Source of Figure 2.11 shown in Materials and Methods].

[202] http://technologyinscience.blogspot.co.uk/2011/09/gel-filtration-chromatography-gf-size.html. [Source of Figure 2.12 shown in Materials and Methods].

[203] F. William Studier. Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification*, 41(1):207–234, 2006.

[204] Geoffrey A. Holdgate, Walter H. J. Ward, and Walter H. J. Ward. Measurements of binding thermodynamics in drug discovery thermodynamics and their use in studies of biomolecular recognition and drug discovery. *Drug Discovery Today*, 10(22):1543–50, 2005.

[205] Remo Perozzo, Gerd Folkers, and Leonardo Scapozza. Thermodynamics of Protein Ligand Interactions: History, Presence, and Future Aspects. *Journal of Receptors and Signal Transduction*, 24(1):1–52, January 2004.

[206] J. J. Christensen, Reed M. Izatt, and D. Eatough. Thermodynamics of Metal Cyanide Coordination. *Inorganic Chemistry*, 4:1278–1280, 1965.

[207] Xingding Zhou, R Manjunatha Kini, and J Sivaraman. Application of isothermal titration calorimetry and column chromatography for identification of biomolecular targets. *Nature protocols*, 6(2):158–65, February 2011.

[208] http://www.biochemistry.ucla.edu/biochem/shared/instruments/itc.html. [Source of Figure 2.15 shown in Materials and Methods].

[209] Ralitza Boubeva. *Understanding tyrosine kinase domain plasticity through identification of protein residues involved in the control of the conformational transition*. PhD thesis, University of Geneve, 2011.

[210] Chung-Jung Tsai, Buyong Ma, and Ruth Nussinov. Folding and binding cascades: shifts in energy landscapes. *Proc. Natl. Acad. Sci. USA*, 96:9970–9972, 1999.

[211] David D Boehr, Ruth Nussinov, and Peter E. Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nature chemical biology*, 5(11):789–96, November 2009.

[212] Michael Deininger, Elisabeth Buchdunger, Brian J Druker, and Washington Dc. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood*, pages 2640–2653, 2005.

[213] Theonie Anastassiadis, Sean W. Deacon, Karthik Devarajan, Haiching Ma, and Jeffrey R. Peterson. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat. Biotechnol.*, 29(11):1039–45, November 2011.

[214] Cai-Hong Yun, Kristen E. Mengwasser, Angela V. Toms, Michele S. Woo, Heidi Greulich, Kwok-Kin Wong, Matthew Meyerson, and Michael J. Eck. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc. Natl. Acad. Sci. USA*, 105(6):2070–5, February 2008.

[215] Mohammad Azam, Markus A. Seeliger, Nathanael S. Gray, John Kuriyan, and George Q. Daley. Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nature structural & molecular biology*, 15(10):1109–18, 2008.

[216] O. Miyashita, J. N. Onuchic, and P. G. Wolynes. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc. Natl. Acad. Sci. USA*, 100(22):12570–12575, October 2003.

[217] Li-Jun Yang, Jun Zou, Huan-Zhang Xie, Lin-Li Li, Yu-Quan Wei, and Sheng-Yong Yang. Steered molecular dynamics simulations reveal the likelier dissociation pathway of imatinib from its targeting kinases c-Kit and Abl. *PloS one*, 4(12):1–8, January 2009.