UNIVERSITÉ DE GENÈVE
Section des Sciences Pharmaceutiques

UNIVERSITÁ DEGLI STUDI DI PADOVA
Department of Pharmaceutical Sciences

FACULTÉ DES SCIENCES
Professeur Leonardo Scapozza

Professeur Stefano Moro

Exploring Structure and Plasticity of Tyrosine Kinase Domains for Drug Discovery

THÈSE
en co-tutelle

présentée à la Faculté des sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences, mention sciences pharmaceutiques

par

Loris MORETTI

de

Cartura (Italie)

Thèse N° 3904

GENÈVE

Atelier d'impression de la Section de Physique

2007

UNIVERSITÉ DE GENÈVE

FACULTÉ DES SCIENCES

Doctorat ès sciences
mention sciences pharmaceutiques

Thèse en cotutelle avec **l'Universita Degli Studi di Padova – Italie**

Thèse de *Monsieur Loris MORETTI*

intitulée :

"Exploring Structure and Plasticity of Tyrosine Kinase Domains for Drug Discovery"

La Faculté des sciences, sur le préavis de Messieurs L. SCAPOZZA, professeur ordinaire et co-directeur de thèse (Unité de Biochimie Pharmaceutique – Laboratoire de chimie thérapeutique), S. MORO, professeur et co-directeur de thèse (University of Padova – Department of Pharmaceutical Sciences – Molecular Modeling Section – Padova, Italy), A. CAVALLI, docteur (University of Bologna – Department of Pharmaceutical Sciences – Bologna, Italy), F. L. GERVASIO, docteur (Eidgenössische Technische Hochschule Zürich – Computational Science – Group Parrinello - c/o USI Campus – Lugano, Suisse), et A. MARSTON, docteur (Laboratoire de pharmacognosie et phytochimie), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 17 octobre 2007

Thèse - 3904 -

Le Doyen, Jean-Marc TRISCONE

N.B.- La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

Nombre d'exemplaires à livrer par colis séparé à la Faculté : - 7 -

per Francesco ...

"When the only tool you own is a hammer,
every problem begins to resemble a nail"

Abraham Maslow

Contents

III

List of Figures

List of Tables

List of Abbreviations

AA	amino acid
ADP	adenosine di-phosphate
ALCL	anaplastic large cell lymphoma
ALL	acute lymphoblastic leukemia
AML	acute myelogenus leukemia
AMP	adenosine mono-phosphate
cAMP	cyclic adenosine mono-phosphate
ATP	adenosine tri-phosphate
BLAST	basic local alignment search tool
CD	circular dichroism
CML	chronic myelogenus leukemia
CPK	Corey, Pauling, Koltun
DNA	deoxyribonucleic acid
ECM	extracellular matrix
FASTA	FAST-all
FDA	food & drug administration
GDP	guanosine di-phosphate
GIST	gastrointestinal stromal tumor
GTP	guanosine tri-phosphate
Hbond	hydrogen bond
HHM	hidden Markov model
ITD	internal tandem duplication
JM	juxtamembrane domain
KD	kinase domain
KID	kinase insert region
MD	molecular dynamics simulation
MM	molecular mechanics
NMR	nuclear magnetic resonance
PCR	polymerase chain reaction
PDB	protein databank
PK	protein kinase
aPK	atypical protein kinase
ePK	eukaryotic protein kinase

PP	protein phosphatase
PSK	protein serine/threonine kinase
PTB	phosphotyrosine binding domain
PTP	protein tyrosine phosphatase
PTK	protein tyrosine kinase
QM	quanto mechanics
RMS	root mean square
RMSD	root mean square deviation
RMSF	root mean square fluctuation
RTK	receptor tyrosine kinase
nRTK	non-receptor tyrosine kinase
SCR	structurally conserved region
SH2	src homology 2 domain
SH3	src homology 3 domain
TK	tyrosine kinase
TM	transmembrane domain
3D	three-dimensional

Summary

The common domain of tyrosine kinase proteins, the kinase domain (KD), is the subject of this thesis. It is the enzymatic moiety that catalyzes the transfer of a phosphoryl group from an ATP molecule to a tyrosine residue of a peptide substrate.

Cells have developed such a device through many evolutionary steps up to the achievement of its stable function and control in multicellular eukaryotes. By domain shufflings and gene duplications, the protein family grew and differentiated in many elements with different topologies. The variety of protein modules at N- and C-terminal position of the KD characterizes TKs for their biological functions and behaviors, interactions with secondary partners and localization within the cell. Two groups of kinases, transmembrane and cytosolic proteins, are present for cell-signaling purposes. While the first ones are able to capture a specific signal from the extracellular environment and transduce it inside the cell, these are called receptor tyrosine kinases. The second proteins are often found in biological pathways to transfer and differentiate various signals inside the cell and are known as non-receptor tyrosine kinases. In humans, 90 tyrosine kinases have been identified and classified in 30 subfamilies based on their sequence similarities.

The crucial physiological role of these proteins in cell growth, development and differentiation has been confirmed by a large body of the scientific literature as well as their correlation with several diseases. Pathologically, many protein modifications, namely mutations, deletions and insertions of stretches of amino acids, are correlated with the abnormal function of TKs. Moreover, protein overexpressions and fusions have been found in cell lines with oncogenic behaviors. Thus, the need to understand the mechanisms and to develop treatments of many diseases where TKs are involved, pushes the scientific community to investigate several aspects of these proteins.

The work here presented goes in such direction. Several features of these evolutionary related enzymatic moieties are evaluated with major focus on their sequences, structures, dynamics and ligand bindings. Issues like: molecular recognition of ligands, stretches of amino acids determining structural characteristics and dynamics of the enzymatic activation were addressed to explore, by means of several computational tools, the kinase domain at the molecular level.

The study started by screening several databases to collect information about the tyrosine kinase domain. Sequences of the proteins were retrieved and investigated by multiple sequence alignments to evaluate the amino acid conservation among the family members. Then, experimentally solved structures were extensively searched into the protein databank and collected. The KD-ligand complexes were compared to find the minimal re-

quirements for a molecule to interact at the ATP-binding site. The defined "ligand-core" identifies the essential structural elements of a ligand binding to TKs. The ligand-core, as presented in chapter 2, is the part of the ligand interacting with the catalytic domain through one hydrogen bond and several hydrophobic contacts with residues whose side chain character is conserved.

Later, the available crystallographic structures were compared to understand three-dimensional conservation as well as variations due to changes at the primary structure level. A classification of KDs was attempted based on the disposition of five clusters of residues observed in different KD states of kinases. In particular, 14 residues of the clusters belonging to the hydrophobic core of the two lobes appeared to be crucial for the structure conformation and dynamics. These are found at the interface between the N- and C-lobes and the residue types and interactions, 2 polar and large hydrophobic contacts, are conserved for the TK subfamilies. These results and further evaluations of the ligand chemistry, conformations and interactions, and the pharmacology features: induced fit versus conformational selection, are described in the third chapter.

The next chapter opens with the evaluation of molecular dynamics simulation as tool to investigate the motion of the kinase domain. Several trajectories were calculated for the non-productive and productive conformations of insulin receptor (Insr) to address the issue of reproducibility of trajectories. The central body of the chapter is dedicated to the exploration of a possible mechanism of the enzyme activation. Flt3 proteins bearing internal tandem duplications (ITDs) at the juxtamembrane (JM) are implicated in some hematopoietic malignancies because the sequence insertion leads to the constitutive activation of the kinase domain. Flt3 motion was computed for the wild-type protein and for several homology models containing ITDs. The hint of these first evaluations, a steric hindrance of the JM acting on the C-helix, was later studied with Flt3 constructs where parts of the juxtamembrane are removed. The freedom achieved by this helix has been identified as the driving force for the N-lobe movements.

The interface residues identified in the third chapter are further investigated in the fifth one. The dynamics of the Insr crystallographic structures of non-productive and productive as well as the one of a third conformation has been investigated with particular focus on the residues at the interface. While the conformation of the majority of these residues appeared to be stable, a rearrangement of interactions for a triad of amino acids at C-terminal position of the C-helix is appreciated as well as conformational changes of the side chain of two residues, one from the hinge and the second from the loop C-terminal to the C-helix.

While performing structure-based drug design *in-silico*, one of the main issues is the selection of a proper candidate as protein structure to target. Often, the available experimental structures are not sufficient to fulfill the need and homology modeling might be used to predict the wanted protein conformation. This method basically works by the identification of structurally conserved regions (SCRs) from a template structure to transpose into the target sequence. The isolation of SCRs should be as accurate as possible to generate a stable and predictive 3D-model. Here, beside the evaluation of Insr structures as possible general templates for homology modeling of tyrosine kinases, those residues

at the interface of lobes are suggested as SCRs.

Concluding, the finding of the ligand-core as common feature of TK complexes can be successfully used as minimal pharmacophore to evaluate KD-ligand interactions especially in virtual screening protocols. Moreover, 14 residues were identified at the interface between the two lobes of the KD and their dynamics has been evaluated for three different conformations of Insr. Indeed, these 14 residues are good candidates among the possible stretches of amino acids involved in the stabilization of differently populated conformations of the KD.
Finally, the activation mechanism for oncogenic and wild type Flt3 was investigated and a major element driving the first step of such conformational change suggested.

Chapter 1

Introduction

1.1 Prologue

1.1.1 The tree of life

Since the first scientists[1], who classified organisms based on their macroscopic character-
istics, the living beings were depicted as subsequent steps of a common evolution. The
discovery of ancient fossils around the world permitted to date and to proof the existence
of animals' ancestors in the Cambrian era[2]. Similarities between different species were
later inferred by the study of their primary components: the cells. Cell cycle, metabolism
and organization in organelles can be compared to find relationships between organisms.

   With early hypotheses and recent experiments our understanding takes a glance on
how life on earth has evolved with incremental complexity and organization. The adapta-
tion to the changing environment is a common aspect for the living beings. To represent
the relation among species, a tree is usually depicted. This symbolic tree is composed
by the trunk, from where the *phyla* of organisms depart as branches, and knots where
the split of two lineages occurred. In the tree of life (examples in figure 1.1 on the
following page) unicellular prokaryote and superior animals share the same provenance
(http://www.tolweb.org).

   The information about our morphological and physiological behavior is "written" in
the genetic material (namely DNA), which is transmitted from parents to children, from
a generation to the next one, from an organism to the next one. The DNA, is composed
by building blocks (nucleotides) and is organized in genes which encode for different
proteins: the function effectors. So, how the diversity of species is achieved? Differences
are attributed to modifications of the genetic material, like mutations of nucleotides and
exchanges, losses and duplications of different gene portions, in order to adapt to the
changing habitat.

---

[1]The first published idea on the matter is "the theory of natural selection" of Charles Darwin (1809-
1882).
[2]Around 540 million years ago, when the explosion of animal diversity seems to have happened.
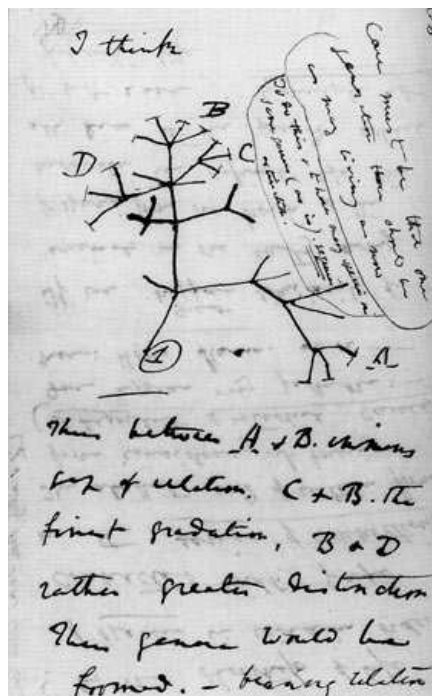
(a) *From Charles Darwin's notes.*
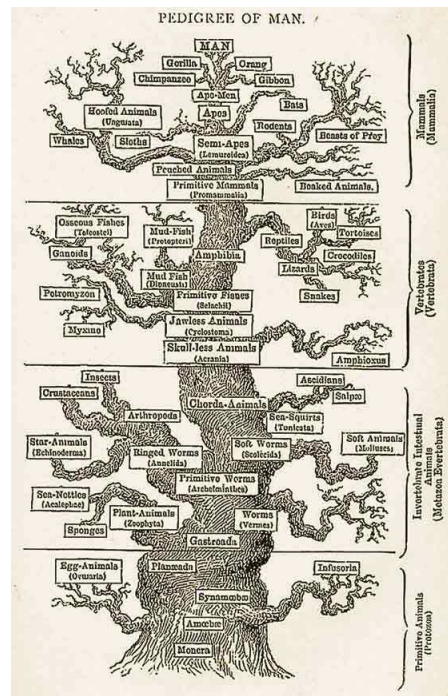
(b) *From the book:* The Evolution of man *by Ernst Haeckel.*

Figure 1.1: *First attempts of tree of life drawings (taken from public domains: www.amnh.org and www.wikipedia.org)*

1.1.2 Phylogenetic studies

In the past few decades, the new knowledges and experimental techniques (i.e. PCR, molecular cloning and sequencing) allowed the phylogenetic analysis at the molecular level. Nowadays, genes are often found as material of many studies because virtually every genes can be cloned, sequenced, expressed and tested. Furthermore, genes can be compared by aligning their nucleotide sequences for matching positions. With the acquisition of genomes (the collection of genes) from different species and the increasing computational power, more sharp border lines between different stages of life's evolution can be drawn by sequence alignments of genes encoding for proteins having the same function in different beings.

These approaches are promising but hide many potential pitfalls such as: i) materials selection, ii) employed methods and iii) results interpretation. The choice of the gene to be aligned is important because natural selection can favor one family over another based on their roles and functions, resulting in different pattern and rate of development [1, 2]. In addition, the used materials can differ for type, nucleotide sequences rather than amino acid sequences of the related proteins, or source, like different databases. One has to be aware of the interdependency that exists between method used, results and interpretation.

1.1.3 Evolution and genes

Due to the continuously upcoming discoveries about new species and their relationships, the tree of life has been redrawn many times in the past and is likely to be in the future. Nowadays, the well accepted division of living beings groups three main kingdoms: eubacteria, archaeabacteria and eukaryotes. Cells of archaeabacteria and eubacteria possess similar morphology but differ in their metabolic pathways displaying different biochemistry. Eukaryotes, differ from the other two reigns for the presence of nuclei and certain organelles, like mithocondria. However, the remarkable feature about the eukaryotes, to whom we belong, is the appearance of multicellularity which later developed in tissue-organized organisms [3].

Although, the large family of eukaryotes has many lineages (figure 1.2 on the next page), our attention is drawn by the macroscopic living beings that widely populate our planet: plants and animals. Thus, many efforts have been spent to investigate the genetic code of these branches which are linked by unicellular common ancestors [2, 5, 6].

Many steps forward have been made by the full genome sequencing of some organisms. In the year 1996 the genome of an yeast *Saccharomyces cerevisiae* [7] was first published, followed by the genomes of a worm *Caenorhabditis elegans* (1998) [8], a plant *Arabidopsis thaliana* (1998) [9, 10], a fly *Drosophila melanogaster* (2000) [11] and the human one, *Homo sapiens* (2001) [12]; later on, many other genomes from different species have been acquired (an up-to-date collection can be be found at "genomic standard consortium", darwin.nox.ac.uk/gsc/gcat).

The development of animals from unicellular ancestors, have been investigated by comparison of genes thought to be main players in their evolution [2, 6, 13–15]. The studied gene families appear to have the major diversification components in two bursts of gene duplications: the first before the parazoan-eumetazoan split, which produced the main subfamilies with different functions, and the second during the early evolution of vertebrate, which caused the enlargement of the subfamilies with new members and tissue-specific expression patterns (in figure 1.2 on the following page both occurs in the animals branch, in violet).

1.1.4 When did tyrosine kinase appear in evolution?

One of the frequently used pools of genes for phylogenetic studies in the animal kingdom is the Protein Tyrosine Kinase family (PTK or TK) [5, 16]. In fact, the proteins translated from these genes have been broadly recognized as implicated in cell signaling, cell-to-cell communications and metabolism [17]. Elements of the family have been widely isolated from multicellular organisms in the eumetazoan reign and only recently the family has been discovered outside of this *phyla* [6, 18, 19].

At this point a specification is needed. The PTKs can be subdivided in 2 groups: Receptor Tyrosine Kinases (RTKs) and non-Receptor Tyrosine Kinases (nRTKs). The two groups are likely to be subsequent stages of the family evolution. First appeared nRTKs and later, by domain shuffling events, the cytoplasmic PTKs could have acquired transmembrane and extracellular domains, as suggested by King and Caroll [19]. Thus, these two groups of genes most likely appeared in the evolution at different time points, first the cytoplasmic elements and later the receptors.
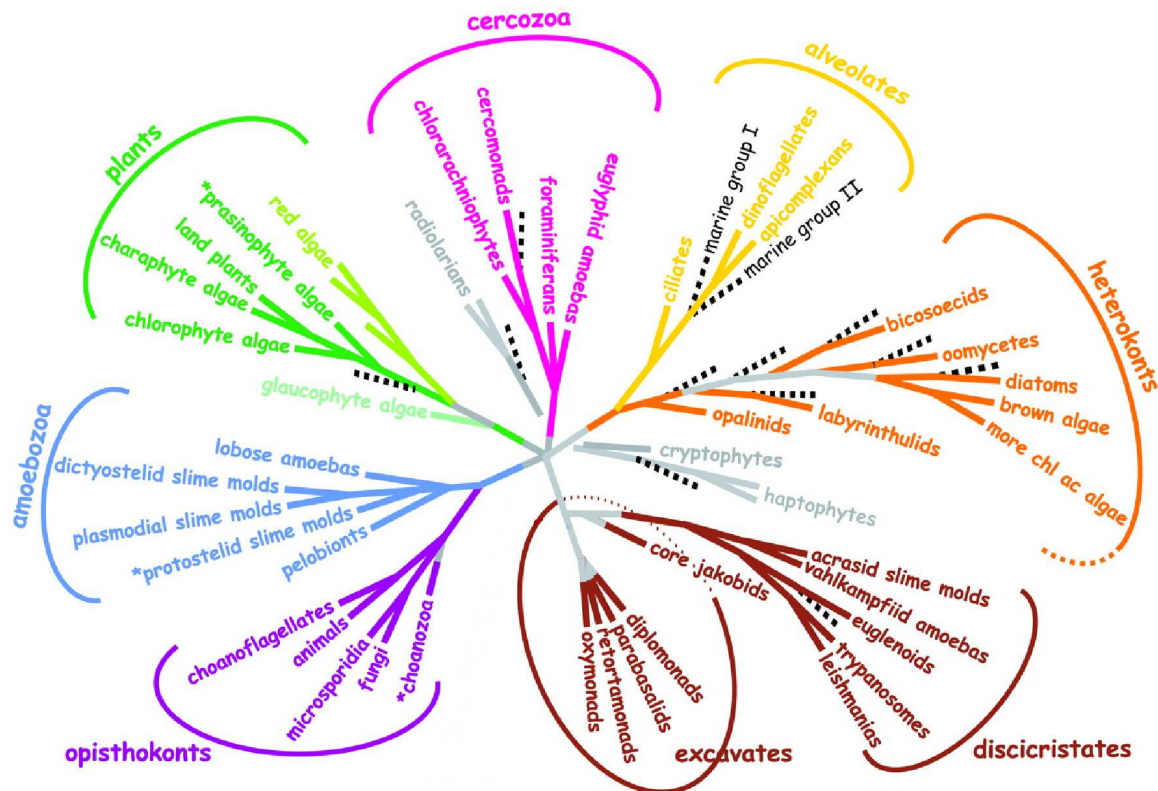
Figure 1.2: *Representation of the tree of the eukaryotes kingdom, adapted from [4]*

Despite the evidence of protein tyrosine kinase activity in yeasts [20], which is probably due to kinases with dual specificity[3] [21], different groups have unsuccessfully screened the yeast genome seeking for PTKs genes [22–24], suggesting the absence of this family in the early eukaryotic organisms.

In the *phyla* of plants, tyrosine phosphorylation events have been also recorded. In cells of *Arabidopsis thaliana*, proteins phosphorylated on tyrosines were immunodetected [25]. This signal could be down regulated by the previous addition of a PTK inhibitor. Another interesting study associated the bending of the petioles of *Mimosa pudica* with a decreased amount of the phosphorylated tyrosines of Actin, a protein of the cytoskeleton [26]. The effect was antagonized by the addition of a inhibitor of Protein Tyrosine Phosphatases (PTPs). Although no tyrosine kinases could yet be isolated from plants and characterized, recently this year, their presence in vegetal species was predicted [27]. Nevertheless the authors revealed also that some of them lack the catalytic residues which characterize the family [27]. Thus, a clear picture about PTKs implants could not yet be drawn. Moreover, the authors of this study also predict the presence of 7 putative TKs in the human intestinal parasite *Entamoeba histolytica* [27].

While there is a report of nRTKs isolated from a slime mold *Dictyostelium discoideum* [18] which showed later tyrosine kinase activity [28], the concomitant isolation

---

[3]Protein kinases which target tyrosine as well as serine and threonine residues of a protein substrate.

of nRTKs and RTKs succeeded only for metazoan organisms. Genes encoding for PTKs were identified in sponges which are multicellular beings close in evolution to eumetazoa [14, 16]. Their multicellular character is revealed, at molecular level, by the presence of Collagen, a protein of the extracellular matrix (ECM), and other adhesion molecules [3]. A more intriguing discovery is the finding of members of tyrosine kinases in choanoflagellates, which are single-cell organisms of aquatic environments [19, 29, 30]. These organisms can spontaneously form organized colonies [31], showing intercellular joined ultrastructures [32]. Cells of choanoflagellates cultivated and tested *in-vitro* have a certain proliferation rate in the presence of nutrients that is significantly reduced when treated with PTKs inhibitors [29].

Segawa et al. compared the activity and regulation of Src-related tyrosine kinases from a sponge and a choanoflagellate [30]. In animals, Src is a cytosolic protein kinase involved in the mediation of a large variety of signals in cells. Src is down-regulated by another PTK: Csk. Although both are present in the two eukaryotes, the negative control appears unstable [30]. The relief of inhibitory mechanisms of Src results in the attenuation of cell-cell interaction [33, 34]. Furthermore, the activity of Csk is higher for the sponge, the first full-time multicellular organism, than for the choanoflagellate, which can be occasionally organized in colonies [30]. Taken together, these findings suggest a central role of tyrosine kinases in the establishment of multicellular communications.

In conclusion, tyrosine phosphorylation has been demonstrated to be an important component of the cell signaling in many stages of the evolution with the detection of its activity even outside of the eukaryotic reign [20, 35]. Although attempts for isolation and characterization of TKs in early species of eukaryotes and outside the *phyla* are needed, a rough picture about their genes evolution can be drawn. These molecular devices might have appeared before choanoflagellates but their tight control and stability are acquired in later stages. In addition, the appearance of tyrosine kinase receptors could be correlated with the acquisition of multicellularity in animals.

1.2 Protein Kinases

Cells respond to external and internal stimuli varying the expression pattern of proteins that lead to a morphological and biochemical change. The term "cell-signaling" is the synthesis of the mechanisms occurring when a stimulus is transduced and transmitted throughout the cell. Protein phosphorylation has been recognized as one of the main player in cell-signaling in eukaryotes with PKs working in a concerted fashion passing the message from one to another in kinase cascades [36]. This behavior is figured out like an electric circuit, with PKs as transistors forwarding the signal from device to device, allowing amplification, feed-back, cross-talk and branching events, till the final target eventually arises the response [37].

These proteins are enzymes that catalyze the phosphoryl-transfer to a peptidic substrate. The reaction is antagonized by the hydrolysis of the phosphate, shutting down the signal flux, by another family of enzymes: the protein phosphatases (PPs). Cells have evolved both families in order to properly control the phosphorylation state of proteins [36]. Despite their role in cell-signaling, knowledge about protein phosphatases is

not as deep as for the counterpart.

In human, approximately 1.7% of the total number of genes is devoted to express PKs [38], representing one of the largest family. These proteins regulate the elementary cell functions participating in metabolism, development, differentiation, movement and apoptosis. In addition, kinases are a major problem when their activity is out of the normal cell control. Deregulation of these enzymes has been largely correlated with pathological states as tumor, diabetes and many others [39].

1.2.1   Protein kinases by dates

Starting from more than half a century ago a large portion of the scientific literature has been devoted to protein kinases, and although many studies surely deserved to be cited, the complete review is not the purpose of this chapter. Thus, to have a taste of the thousands efforts, here, just an overview of the achieved milestones is given.

The first protein phosphorylation reaction catalyzed by an enzyme was published in the 1954 by Burnett and Kennedy [40]. The authors, revealed the presence of a protein in the rat liver mithocondria which selectively transferred radioactive phosphate of ATP to Casein protein. Soon after, the phosphorylase b kinase (cyclic AMP-dependent kinase, Pka) was extracted from rabbit muscle and characterized for pH, Ca^{++} and ATP dependencies by Krebs *et al.,* [41, 42]. In 1979, the regulation of protein activity by reversible phosphorylation with a kinase and a phosphatase was shown [43]. In the same year, the physiological role of the auto-phosphorylation was suggested when two phosphorylation sites were identified on the catalytic subunit of cAMP-dependent kinase by high voltage electrophoresis [44]. Later, peptides competing with the substrate binding to a calmodulin dependent kinase were chemically synthetized [45]. Undoubtedly, a big step forward was taken with the publication of the protocol for protein kinase expression in a host organism, yielding high product amount [46]. Thus, Pka was the first recombinant protein kinase expressed in *Escherichia coli*, purified and later characterized.
Till the year 1979 protein phosphorylation was demonstrated only toward serine and threonine residues but then, enzymatic activity directed to phosphorylated tyrosine residues was detected in polyomavirus [47], Rous sarcoma virus [48] and Abelson leukaemia virus [49]. The transforming activity of these viruses was soon associated with cellular homologue components regulating cell growth [50].
The two main families of protein kinases were discovered and, in the following years, many works focused on the isolation of protein kinases in different organisms. Elements of the protein superfamily were identified, investigated and classified in families based on sequence characteristics [51, 52].
In the year 1991, the first notion of the three-dimensional atomic disposition of the catalytic core of a protein kinase in apo-form and in complex with substrate was acquired by solving the structure of cAMP-dependent kinase by means of X-ray crystallography [53, 54].
Other protein portions of the protein kinases have been later recognized as implicated in the signal propagation with a modular fashion [55]. With the publication "The protein

kinase complement of the human genome" [38], the human "kinome" [4] have been finally defined in numbers and constitution.

1.2.2 Classification of protein kinases

The current index of protein kinases is based on the similarity of the amino acid sequence of their catalytic domain, the kinase domain (KD). The early identified sequence features to sort out protein kinases [52] are still widely used for classification purposes. Recently, based on hidden Markov models (HMMs), genomes have been screened to update the kinome classification [27, 38].

In human, the first division is in two groups: the "conventional" eukaryotic protein kinases (PKs), with 478 members, and the "atypical" protein kinases (aPKs), with 40 members. The aPKs are proteins reported to have biochemical kinase activity [38]. Despite the very low sequence similarity in the catalytic domain (even below 10%) some, namely alpha and RIO families, have structural similarities to the eukaryotic protein kinases [56–58] (see figure 1.4 on page 12).

The PK superfamily is subdivided in 9 groups:

1. the **AGC family** includes the cAMP-dependent protein kinases (Pka), protein kinase G and protein kinase C and is composed by 63 members;

2. the **CAMK family** includes the Ca^{++}/Calmodulin-dependent kinases and has 74 members;

3. the **CK1 family** includes the casein kinase 1, related kinases and is formed by 12 members;

4. the **CMGC family** includes the cyclin-dependent kinases, mitogen-activated kinases, Cdk-like kinases, glycogen synthase kinase and is composed by 61 members;

5. the **RGC family** includes receptor guanylate cyclase kinases and has 5 members;

6. the **STE family** includes Ste20 kinases and is formed by 47 members;

7. the **TK family** includes tyrosine kinases and is composed by 90 members;

8. the **TKL family** includes tyrosine kinase-like kinases and has 43 members;

9. **Others** are kinases which do not belong to the previous families and are 83.

This classification has recently been slightly modified by a study where a multi-level HMM library of protein kinases has been developed and applied to 21 eukaryotic genomes [27]. The authors were able to reclassify the "**others**" group of PKs into the 8 known PKs families and redefine the whole set of PK groups as: 84 ACGs, 98 CAMKs, 12 CK1s, 70 CMGCs, 5 RCGs, 61 STEs, 93 TKs and 55 TKLs [27].

---

[4]The whole ensemble of the genes transducing for kinases, in a certain species

1.3 Kinase Domain

1.3.1 The sequence

The common element of protein kinases, the catalytic domain, is composed by a number of amino acids ranging from 250 to 300. The amino acidic sequence of kinase domains shares certain similarities in different positions. The conserved characteristics which define kinases as homologues can be extrapolated from the multiple sequence alignment of the primary structures of diverse PKs (figure 1.3.1 on the facing page). Historically, the definition of these common features developed till the depiction of 12 subdomains (roman numbers in figure 1.3.1 on the next page) [51, 59]. Reasons of the conservation were revealed when the three-dimensional structure of the catalytic domain became available (figure 1.6 on page 14). The rational is that an identical or similar residue, at certain position in the sequence and structure possesses the same function:

- *Flexibility.* Subdomain I, with the consensus GxGxaG[5] (where "x" can be any residue and "a" a hydrophobic one) folding in the loop between the first two strands.

- *Secondary structure maintenance.* Subdomains IV, V, XII and X, have a structural role. The latter possesses a nearly invariant aspartate that fixes the position of residues at subdomain VII.

- *Substrate binding.* Subdomain VII, with the consensus HRDLKxxN for protein serine/threonine kinases and HRDLAARN or in alternative HRDLRAAN (in the case of Src subfamily) for protein tyrosine kinase is the stretch where the two groups of kinases can be differentiated.

- *Catalytic involvement and cofactor binding.* Subdomain II contains the consensus VAVK, with alanine interacting via van der Waals contact with the adenine moiety of ATP and lysine interacting with the phosphate. Subdomain VIII with the consensus DFGL, where the aspartate binds the counter-ion of ATP. Subdomain V is the so-called hinge region with a residue forming a hydrogen bond to the adenine moiety via the backbone.

- *Catalytic residue positioning.* Subdomain III with a nearly conserved glutamate residue which helps the positioning of lysine of subdomain II.

- *Substrate recognition.* Subdomain IX.

1.3.2 The structure

The three-dimensional disposition of the atoms of the kinase domain was revealed in 1991 with the structures of cAMP-dependent kinase [53, 54]. Later on, the collection of solved crystal structure of PKs has grown so large that it becomes difficult today to keep track

---

[5]From the multiple sequence alignment including the RGC kinase sequence (not found any report in the literature) this sequence is elongated by one residue, GxGxxxG. The other four members of this family does not appear to have this consensus when aligned (data not show).

1st β-strand   2nd β-strand   3rd β-strand   1st α-helix

P-loop   C-helix

```
IR....(TK)    1023  ITLLRELGQGS.FGMVYEGNARDIIKG....EAETRVAVKTVNESA......SLRERIEFLNEASVMKG..FTCH
PKACα.(AGC)   44    FERIKTLGTGS.FGRVMLVKHKET.........GNHYAMKILDKQKVV....KLKQIEHTLNEKRILQA..VNFP
CaMK1α(CAMK)  20    YDFRDVLGTGA.FSEVILAEDKRT.......QKLVAIKCIAKEALEG......KEGSMENEIAVLHK..IKHP
CK1α..(CK1)   17    YKLVRKIGSGS.FGDIYLAINITN.......GEEVAVKLESQ.........KARHPQLLYESKLYKI.LQGGV
CDC2..(CMGC)  4     YTKIEKIGEGT.YGVVYKGRHKTT.......GQVVAMKKIRLESEEEG.....VPSTAIREISLLKE..LRHP
MAP3K2(STE)   356   WRLGKLLGQGA.FGRVYLCYDVDT.......GRELAVKQVQFDPDSPETS..KEVNALECEIQLLKN..LLHE
TGFαR1(TKL)   205   IVLQESIGKGR.FGEVWRGKWR.........GEEVAVKIFSS.........REERSWFREAEIYQTVMLRHE
ANPa..(RGC)   528   GSRLTLSGRGSNYGSLLTTEGQFQVFAKTAYYKGNLVAVKRVNRK.......RIELTRKVLFELKHMRD..VQNE
```

G G   fg v   vAvK   E

I   II   III   IV

4th β-strand   5th β-strand   2nd α-helix   3rd α-helix

N-Lobe <hinge> C-Lobe

```
IR....(TK)    1085  HVVRLLGVVSKGQ....PTLVVMELMAHGDLKSYLRSLRPEAENNPGRPPPTLQEMIQMAAEIADGMAYLNA...
PKACα.(AGC)   103   FLVKLEFSFKDNS....NLYMVMEYVPGGEMFSHLRRIG.........RFSEPHARFYAAQIVLTFEYLHS...
CaMK1α(CAMK)  77    NIVALDDIYESGG....HLYLIMQLVSGGELFDRIVEKG.........FYTERDASRLIFQVLDAVKYLHD...
CK1α..(CK1)   72    GIPHIRWYGQEKD....YNVLVMDLLGP.SLEDLFNFCSR.........RFTMKTVLMLADQMISRIEYVHT...
CDC2..(CMGC)  62    NIVSLQDVLMQDS....RLYLIFEFLSM.DLKKYLDSIPP.........GQYMDSSLVKSYLYQILQGIVFCHS...
MAP3K2(STE)   417   RIVQYYGCLRDPQE..KTLSIFMEYMPGGSIKDQLKAYG.........ALTENVTRKYTRQILEGVHYLHS...
TGFαR1(TKL)   258   NILGFIAADNKDNGTWTQLWLVSDYHEHGSLFDYLNRYT.........VTVEGMIKLALSTASGLAHLHMEIV
ANPa..(RGC)   594   HLTRFVGACTDPP....NICILTEYCPRGSLQDILEN........ESITLDWMFRYSLTNDIVKGMLFLHNG..
```

i   l   l   y l h

IV   V   VI

6th β-strand   7th β-strand   4th α-helix

C-loop   A-loop

```
IR....(TK)    1153  .....KKFVHRDLAARNCMVAH...DFTVKIGDFGMTRDIY.....ETDYYRKGGKGLLPVRWMAPESLKDGVF.
PKACα.(AGC)   161   .....LDLIYRDLKPENLLIDQ...QGYIQVTDFGFAKRVKG.........RTWTLCGTPEYLAPEIILSKGY.
CaMK1α(CAMK)  135   .....LGIVHRDLKPENLLYYSLDEDSKIMISDFGLSKMED.........PGSVLSTACGTPGYVAPEVLAQKPY.
CK1α..(CK1)   130   .....KNFIHRDIKPDNFLMGIGRHCNKLFLIDFGLAKKYRDNRTRQHIPYREDKNLTGTARYASINAHLGIEQ.
CDC2..(CMGC)  122   .....RRVLHRDLKPQNLLIDD...KGTIKLADFGLARAFG.......IPIRVYTHEVVTLWYRSPEVLLGSARY
MAP3K2(STE)   477   .....NMIVHRDIKGANILRDS..TGNVKLGDFGASKRLQT....ICLSGTGMKSVTGTPYWMSPEVISGEGY.
TGFαR1(TKL)   322   GTQGKPAIAHRDLKSKNILVKK...NGTCCIADLGLAVRHDSA...TDTIDIAPNHRVGTKRYMAPEVLDDSINM
ANPa..(RGC)   654   .....AICSHGNLKSSNCVVDG...RFVLKITDYGLESFRDLD.......PEQGHTVYAKKLWTAPELLRMASPP
```

hrdlk N l   DfGl   ape

VII   VIII   IX

```
                                    5th α-helix                                6th α-helix
                                  ooooooooooooooooo                          ooooooooo

Insr....(TK)    1214   .....TTSSDMWSFGVVLWEITS..........LAEQPYQGL.......SNEQVLKFVMDGG..............
Pkacα.(AGC)      217   .....NKAVDWWALGVLIYEMAAG.........YPPFFAD.......QPIQIYEKIVS................
Camk1α(CAMK)     196   ....SKAVDCWSIGVIAYILLCG.........YPPFYDE.......NDAKLFEQILKAEYEFDSPYWDDISDS
Ck1α..(CK1)      199   .....SRRDDMESLGYVLMYFNRT........SLPWQGLKAATKKQKYEKISEKKMSTP.....
Cdc2..(CMGC)     182   .....STPVDIWSIGTIFAELATK.........KPLFHGD.......SEIDQLFRIFRALGTPNNEVWPEVESL
Map3k2(STE)      539   .....GRKADIWSVACTVVEMLTE.........KPPWAEF.......EAMAAIFKIATQPTN.............
Tgrαr1(TKL)      391   KHFESFKRADIYAMGLVFWEIARRCSIGGIHEDYQLPYYDLVPSDP..SVEEMRKVVCEQKLR............
Anpa..(RGC)      714   VRG..SQAGDVYSFGIILQEIAL.........RSGVFHVEGLDL....SPKEIIERVTRG..............


                             D       g
                        _____       _____
                                        X                                            XI
```

```
                                    7th α-helix
                                  ooooooooooo

Insr....(TK)    1254   ..........................YLDQPDNCPERVTDLMRMCWQFNPKMRP.....TFLEIVNLL
Pkacα.(AGC)      254   ..........................GKVRFPSHFSSDLKDLLRNLLQVDLTKRFGNLKNGVNDIKNHK
Camk1α(CAMK)     249   AKDFIRHLMEKDPEKRFTCEQALQHPWIAGDTALDKNIHQSVSEQIKKNFAKSKWKQAFNATAVVRHM
Ck1α..(CK1)      245   ..........................VEVLCKGFPAEFAMYLNYCRGLRFEEAP.....DYMYLRQLF
Cdc2..(CMGC)     235   QDYKN...........TFPKWKPGSLASHVKNLDENGLDLLSKMLIYDPAKRI.....SGKMALNHP
Map3k2(STE)      580   ..........................PKLPPHVSDYTRDFLKRIFVEAKLRP.....SADELLRHM
Tgrαr1(TKL)      452   ..................PNIPNRWQSCEALRVMAKIMRECWYANGAARL.....TALRIKKTL
Anpa..(RGC)      759   ..................EQPPFRPSLALQSHLEELGLLMQRCWAEDPQERP.....PFQQIRLTL


                        _____
                                              XII
```
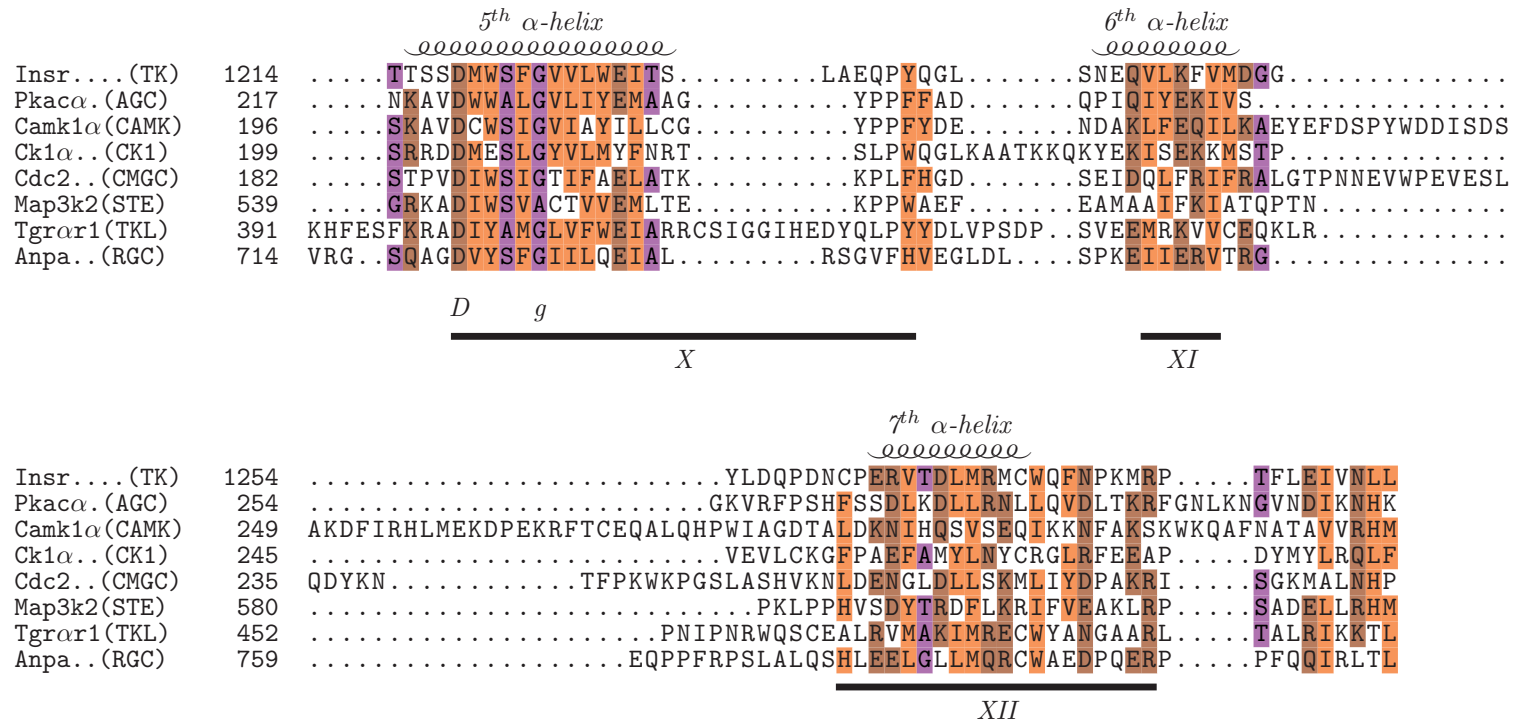
Figure 1.3: *Multiple sequence alignment of eukaryotic PKs coupled with structural alignment. The sequence of Anpa (RGC kinase whose structure is yet unknown) was implemented by means of the program CLUSTALX [60] in the structural alignment of the representative of the other seven families previously produced with BODIL [61]. The alignment was rendered with $T_EX$shade [62]. Names and Numbers concerning each protein sequence are shown on the left. From the top in the first line the label of the secondary structures (drawn in symbols in the second line) referred to the Insr. In the third line, the protein portions (P-loop, C-helix, hinge, C-loop and A-loop) referred to the figure 1.6 on page 14 are depicted. Below the sequences, the main consensus adapted from [52] with highly (capital letters) and well (small letters) conserved residues is shown. At the bottom the twelve subdomain according to [52] are listed. Residues at position where a major character is conserved are color coded for polar (light brown) apolar (violet) and small side chain, namely alanine, glycine, threonine an serine (orchid). All sequences were taken from EXPASY database under the entries are P06213 for Insr, P17612 for Pkacα, Q14012 Camk1α, P48729 Ck1α, P06493 for Cdc2, Q9Y2U5 for Map3k2, P36897 for Tgfαr1 and P16066 for Anpa.*

of the complete set.

The comparison of KD from different kinase families is intriguing (figure 1.4 on the next page); the conservations of the folding and the secondary structures are greater compared to the primary structure (in some cases below 20% of identity). The structural similarity with the "atypical" kinase [6] of the alpha family (figure 1.4h on the following page) is astonishing. The correspondence of some structural elements (mainly in the N-lobe and a helix in the C-lobe) led scientists to hypothesize a common protein ancestor [58, 63].
It is very likely that the high sequence variability among kinases, achieved throughout their evolution, characterizes the protein behaviors (like substrate recognition, inhibition and dynamics) in order to differentiate the same chemical mechanism for a large variety of functions in cell physiology.

The primary structure of catalytic domain folds in an independent macromolecule of about 30 Kilo Dalton, which is a soluble protein found in the cytoplasmatic environment. According to the knowledge about the catalytic domain, the primary biophysical/biochemical functionalities are:

1. binding and orientation of the complex formed by cofactor and counter-ion/s. Mainly ATP but also GTP, with phosphate groups coordinated by one or two divalent cations such as Mg^{++} or Mn^{++};

2. binding and orientation of the peptide substrate;

3. transfer of the γ-phosphate of the cofactor to a side chain hydroxyl group of a residue (either serine or threonine or tyrosine) in the substrate.

The classic three-dimensional arrangement of this domain, depicted in figure 1.5 on page 13, is bilobate. The two substructures are called N-lobe and C-lobe, and connected through a segment: the hinge region.

N-lobe

The amino-terminal part of the domain forms the N-lobe, which is smaller then the C-lobe. It includes a main β-sheet core, composed by 5 antiparallel strands, and a α-helix (α1) (subdomain III see figure 1.3.1 on page 9 and figure 1.5 on page 13), called C-helix spanning between β-strands 3 and 4. The amino acid stretch between the C-helix and the fourth strand folds in a loop which largely interacts with the C-lobe. The loop in between the first two β-strands (β1 and β2), corresponding to subdomain I (see section 1.3.1 on page 8), interacts with the cofactor (figure 1.7 on page 16). This element is highly flexible because of the presence of glycines and can adapt against the α and β phosphates of the cofactor. The alternative names of this flexible element are the Gly-rich loop, P-loop or nucleotide-binding loop. The critical role of this loop appeared to be the cofactor γ-phosphate localization via polar interaction with the backbone amide of the residue after the second glycine in the consensus [65].
The third strand (β3) contains the second subdomain with a well conserved alanine, that

---

[6]Because of the topic of the thesis, the aPKs will be no longer discussed.

(a) *AGC family*

(b) *CAMK family*

(c) *CK1 family*

(d) *CMGC family*

(e) *STE family*

(f) *TKL family*

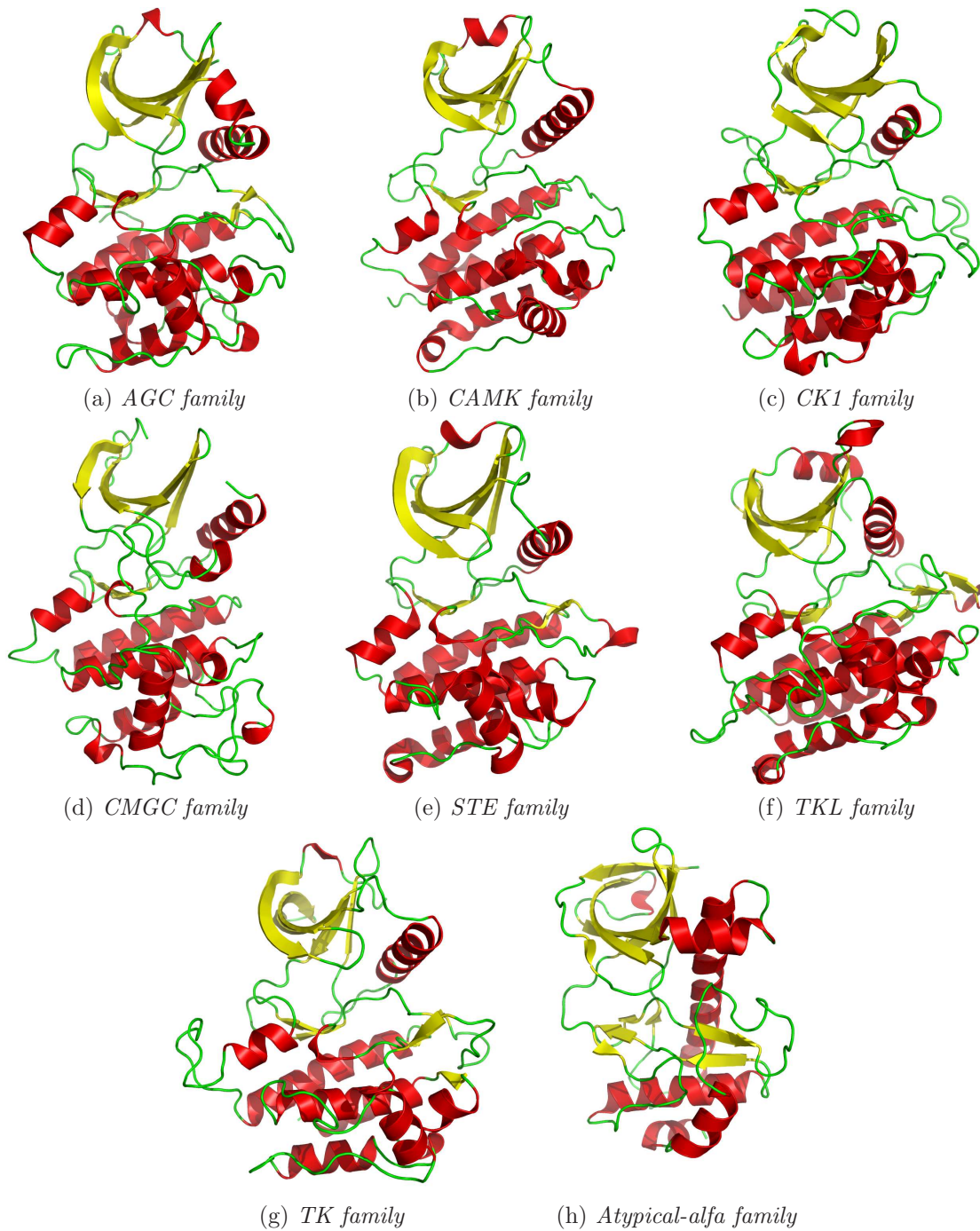(g) *TK family*

(h) *Atypical-alfa family*

Figure 1.4: *Cartoons of representative 3D-structures of the PK families. Secondary structure is represented and colored red (helices), yellow (strands) and green (coils). The three-dimensional coordinates, PDB codes in order of appearance: 2CPK, 2BDW, 1CSN, 1B38, 2HY8, 1IAS, 1IR3 and 1IA9, were rendered with* PYMOL *[64].*
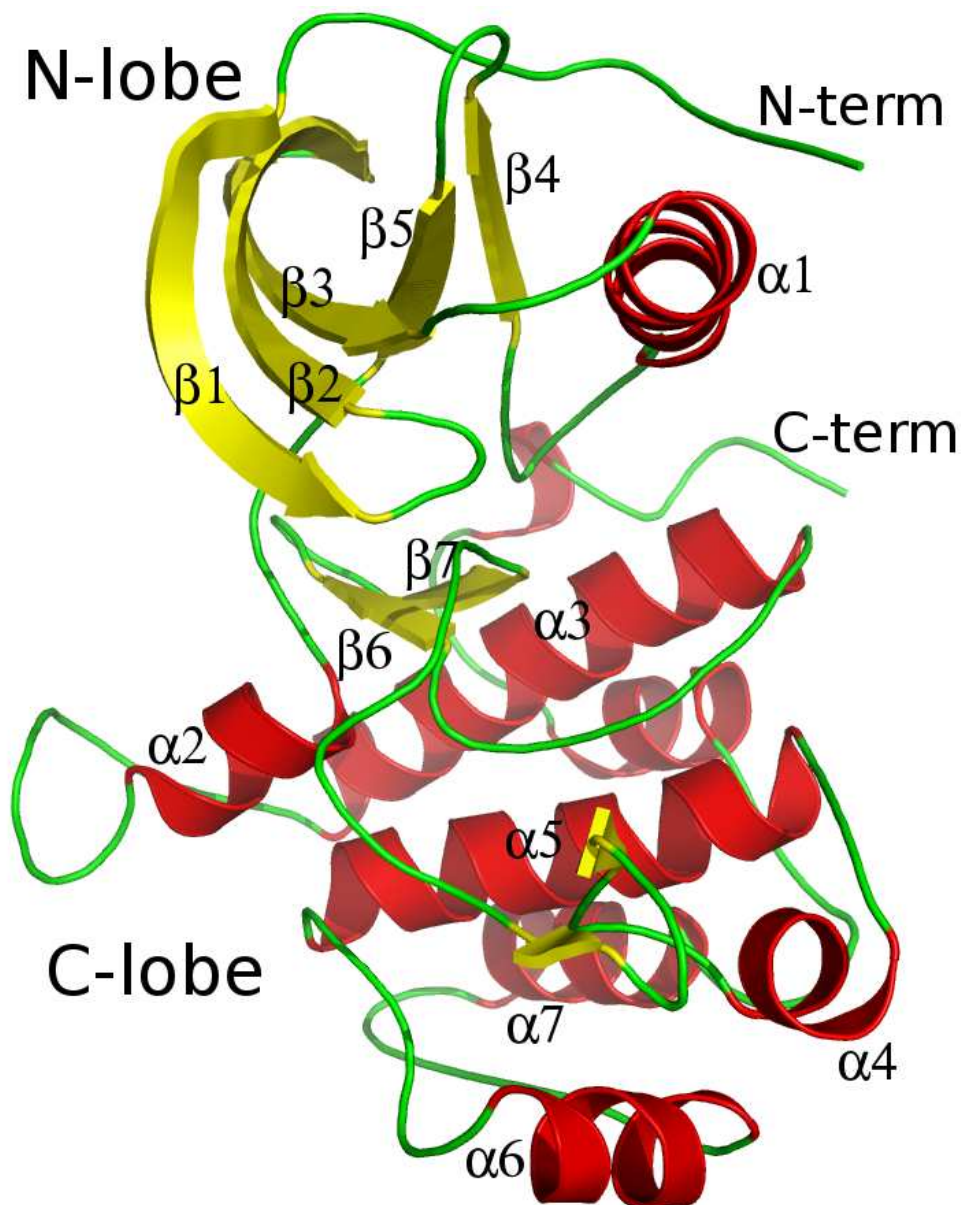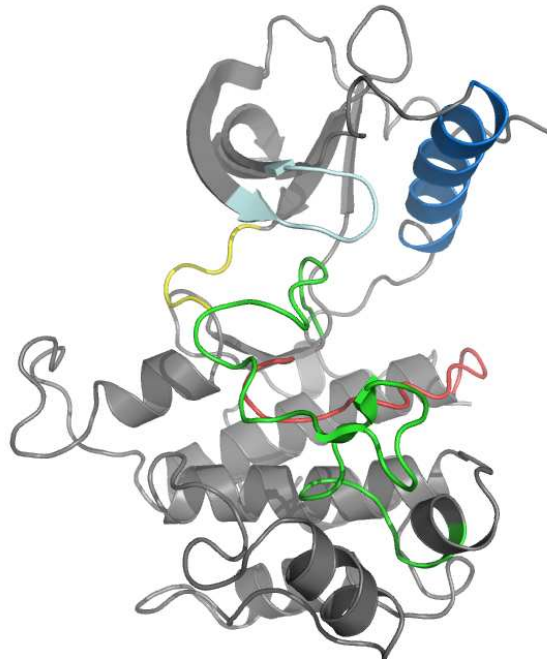
Figure 1.5: *Three-dimensional arrangement of the catalytic domain of TKs. The secondary structures elements of the crystallographic structure of Insr (PDB code: 1IRK) are rendered with cartoon representation and colored: yellow for strands, green for coils and red for helices. Atomic coordinates were drawn with the program* PYMOL *[64].*

(a) *Inactive conformation of the KD of the Insulin Receptor Kinase*
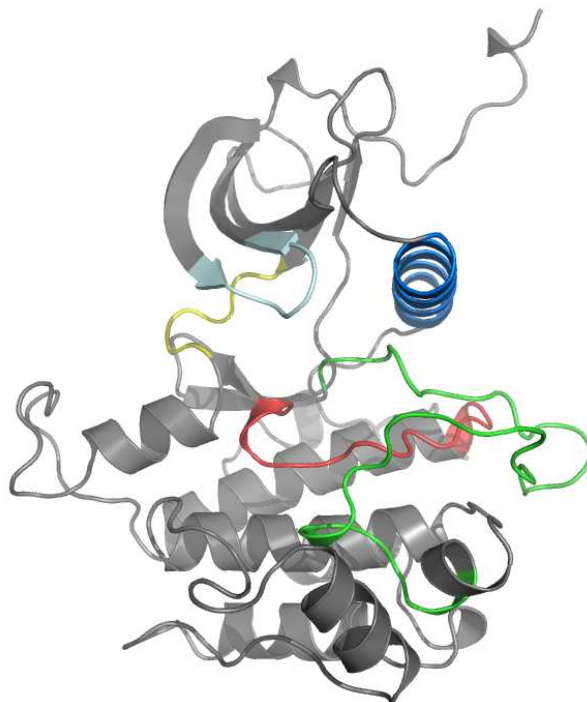
(b) *Active conformation of the KD of the Insulin Receptor Kinase*

Figure 1.6: *Cartoons' representations of crystallographic structures of the inactive state (PDB code: 1IRK) and the active (PDB code: 1IR3) of the Insulin Receptor (a,b). The secondary structure of the protein is in grey with Gly-rich loop (cyan), C-helix (marine blue), hinge region (yellow), catalytic loop (C-loop, in red) and activation loop (A-loop in green). Atomic coordinates rendered with the program* PYMOL *[64].*

interacts with the ring system of the adenine of the ligand, and the conserved lysine coordinating the phosphates. The last two β-strands ($\beta 4$ and $\beta 5$) of this lobe are mainly hydrophobic.

C-lobe

The biggest of the two protein substructures has a helical core composed by six conserved α-helices (figure 1.5 on page 13). From the C-terminus of $\alpha 3$ to the first β-strand ($\beta 6$) there is a loop which is involved in the reaction catalysis (the catalytic loop or C-loop; in red in figure 1.6 on the preceding page). This sequence forms the seventh subdomain (see figure 1.3.1 on page 9) whose residues are involved in the substrate binding and in the phosphorylation reaction. The loop is anchored to the C-lobe core via conserved interaction with an aspartate from the third α-helix ($\alpha 3$).
The two subsequent short strands ($\beta 6$ and $\beta 7$; three amino acids each one) are forming the β-sheet facing the small lobe. Amino acids of these elements are interacting with the cofactor.
From the latter strand a loop is emerging; this is a long loop (with variable size for the superfamily proteins), called activation loop or T-loop or A-loop. This element usually contains one residue which can mimic the substrate (auto-inhibitory role) and when phosphorylated fully activates the enzyme. The main characteristic of this loop is its own mobility: "It is a multipurpose structural element with highly variable structural behavior" [66]. In fact, it can fluctuates among several conformations interfering with the binding of either ATP and/or the substrate. The conformation of this element can be stabilized, upon phosphorylation or not, in a open position as depicted in figure 1.6a or in a close conformation (figure 1.6b on the preceding page). The residues at the A-loop carboxy-terminal are involved in substrate recognition and binding. The activation loop sequence (see figure 1.3.1 on page 9) starts and ends with two quite conserved sequences: the DFG motif (subdomain VIII) and the APE motif (subdomain IX), respectively.
The DFG motif takes part to the chemical reaction through its aspartate and can interfere with the ATP binding depending on the adopted conformation:

Asp-out The aspartic residue points out of the reaction center, while the phenylalanine is in hydrophobic contact with residues at the ATP binding site (see figure 1.7a on the next page). This orientation has been depicted in several but not all PKs.

Asp-in The aspartate is pointing in the reaction center, ready to interact with the Magnesium coordinating β and γ-phosphate groups of the cofactor (figure 1.7 on the following page). The phenylalanine packs between the N-lobe residues, from the C-helix, the 4th strand, the loop in between, and the third strand, and the C-lobe with residues from the catalytic loop and the third helix (see figure 1.7b on the next page).

The rest of the protein is mainly α-helical until the carboxy-terminus. It comprises the last three subdomains (see section 1.3.1 on page 8) whose function is mainly structural taking part of the hydrophobic core of the C-lobe.
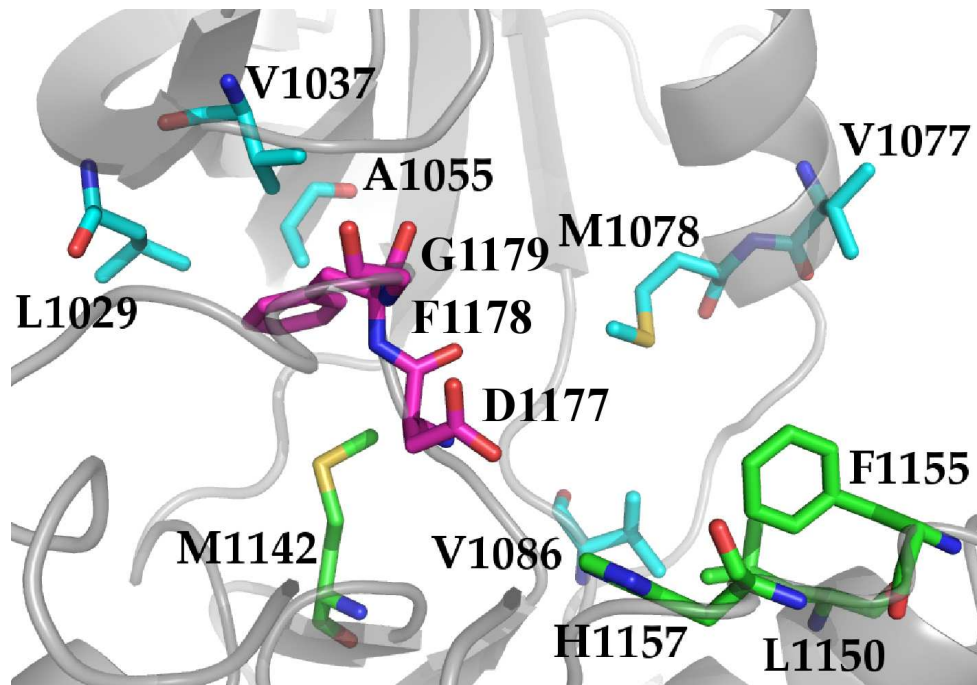
(a) *DFG with Asp-out conformation (representation of the PDB: 1IRK)*
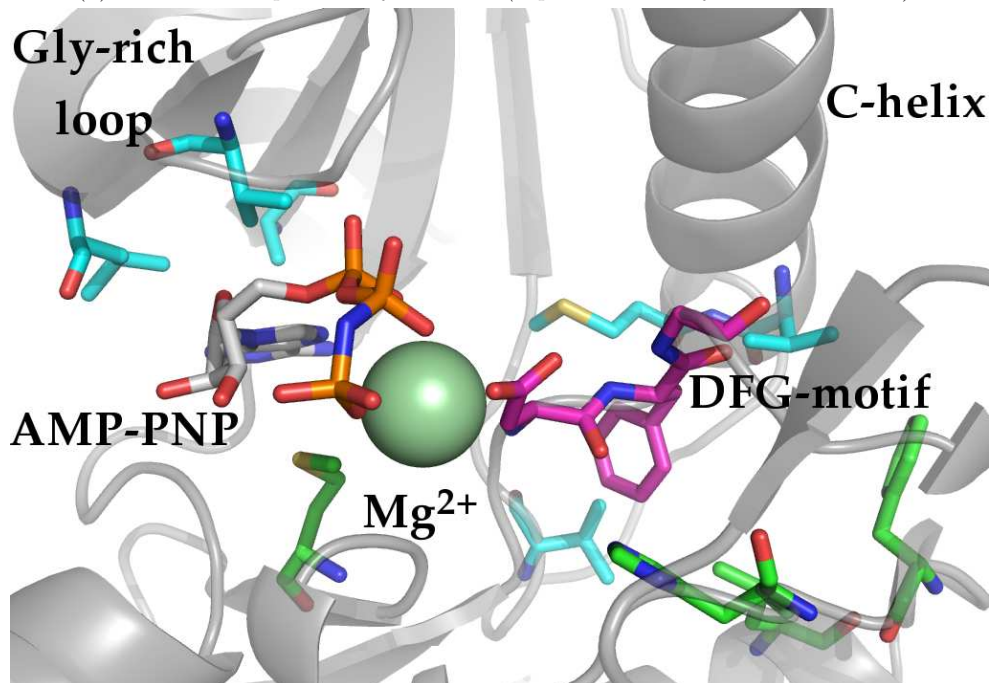
(b) *DFG with Asp-in conformation (representation of the PDB: 1IR3)*

Figure 1.7: *Conformational change of the DFG-motif of the insulin receptor. The secondary structure of the KD is in grey; the residues involved in catalysis and ligand-binding, and the ligand are shown in sticks representations with the atoms colored nitrogen (blue), oxygen (red), sulfur (yellow), magnesium (pale green) and carbon (cyan N-lobe residues, green C-lobe residues, magenta DFG-motif residues and gray AMP-PNP and labeled in (a)). Characteristic elements are labeled in (b). Atomic coordinates rendered with the program* PYMOL *[64]*

Protein states

Concerning the reaction catalysis, from a biochemical point of view, kinases can be detected in two different states, either productive, when the protein is able to bind both substrate and cofactor and complete the reaction of transfer, or non-productive, with the protein auto-inhibited not fully prone for catalysis. While, the productive state can be identified with a well defined structure, the non-productive one cannot. In fact, the latter state can be an ensemble of conformations.

Anyway, two protein conformations are defined: the active and inactive one, representing the two borders of the conformational range where the KD structure can be found. Based on the insulin receptor tyrosine kinase KD, whose two states could be photographed by means of X-rays [67,68], the two forms are depicted (see figure 1.6 on page 14). Actually, it is not always the case because different KDs appear to have different motions as suggested by different inactive states and kinetics of catalysis [69,70].

To make things even more complex, there are other domains and subdomains in the same or different polypeptide chain which can regulate the enzyme activity. Enhancements are seen in the case of Cdk [71] and Egfr [72] where the intervention of a partner is required in order to be fully productive. On the other hand, decrements are achieved in the case of Insr and Pdgfr subfamily members by trapping the kinase in the inhibited form [67,73] (see subsection 1.3.5 on page 26).

Ligand binding

As stated previously, the kinase domain's main functions are the binding and orientation of the reagents for the phosphoryl-transfer reaction. The subsequent description of the bindings is based on the insulin receptor crystallographic structure deposited with the PDB code 1IR3.

The protein is in complex with a short substrate and a ligand, AMP-PNP, which has an oxygen replaced by a nitrogen atom compared to ATP and is supposed to interact in a similar manner as the cofactor [68]. In the active state, the kinase domain binds the cofactor at the cleft, called ATP-binding site, at the interface of the two lobes as depicted in figure 1.8 on page 19. Based on the mode of interaction, the complex protein-ligand can be subdivided in two regions: the first with many van der Waals contacts (figure 1.8b on page 19) and a second one, where mainly polar bonds are found (figure 1.8c on page 19). The adenine ring system is hydrophobically sandwiched by several residues of the two β-sheets (see figure 1.8b on page 19) and forms two hydrogen bonds (H-bond) to the hinge region. One H-bond with the carbonyl group of Glu1104 and the second with the nitrogen atom of Met1106 of insulin receptor as reference structure (figure 1.8b on page 19). The rest of the ATP molecule, namely the sugar moiety and phosphate groups, interacts mainly with polar residues (figure 1.8c on page 19). Asp1110 forms H-bonds with the 3' and 2' -OH groups of the ribose. The negative charges of the three phosphate groups is counter-balanced usually by the Lys1057 in Insr, the conserved lysine of the second subdomain (figure 1.8c on page 19), and by two magnesium ions, which are likely to be coordinated by the ligand before entering the binding site. The two ions are further chelated by water molecules and two conserved protein residues: aspartate of the DFG-motif and the asparagine of the VII subdomain, at positions 1177 and 1164, respectively,

in insulin receptor (in figure 1.3.1).

Concerning the ATP-binding site, Vulpetti and Bosotti carried out a comparison of some KD structures together with a multiple sequence alignment of all protein kinases [74]. The authors defined the pocket by a cluster of 38 residues, around the bound molecule, as well as their conservation and variability throughout the superfamily.

With a smaller dataset, comprising only tyrosine kinases, we have described a main core for ligands and five protein residues interacting with it as the minimal requirements for binding at this pocket [75].

The large number of protein kinases in mammals (more than 500 in humans) and the amount of cellular functions concerted by these enzymes, tell us that the specificity of the reaction target is required to control their activity among a myriad of possible substrates. Kinase domains present different specificity toward the sequence substrate to bind and phosphorylate. The substrate, which can range from a small peptide to a large protein, interacts with the protein at the surface of the C-lobe in correspondence of the C-terminus of the A-loop (see figure 1.8 on the facing page). In fact, the A-loop plays a crucial role for substrate recognition and for creating the environment for the catalytic residues [76]. Substrate specificity is achieved in two steps. The first is related to the residue whose side chain hydroxyl group is going to be phosphorylated. This amino acid, conventionally identified at the P0 position of the substrate, can be either serine or threonine, phosphorylated by protein serine/threonine, or tyrosine, involving protein tyrosine kinases instead. The secondary recognition for substrate's selection is based on the neighborhoods of P0. Residues found at its N-terminal are called P-1, P-2 etc. and those at the C-terminal called P+1, P+2, etc. (figure 1.8a on the next page). The determinants for P0 specificity lie mainly in the catalytic loop, as reported in section 1.3.1 on page 8. Furthermore other positions can be selected via physical/chemical complementarity for residues at the protein surface. Another important interaction can be due to the formation of a short stretch of antiparallel β-sheet between residues from P+1 to P+3 and a portion of the A-loop [67].

1.3.3 The dynamics

The complete conformational energy landscape of the kinase domain, will probably look tremendously complicated, with energy hills and holes, and different from one kinase to another. This protein folding appears to be highly flexible visiting many conformations and, the thermodynamic equilibria can be perturbed toward fewer possibilities by several physiological or pathological mechanisms. Residues phosphorylation, ligand binding, protein-protein interactions as well as residue or sequence mutations are likely to affect the general stability of the catalytic domain.

Thus, one of the major problems studying kinase domains, as well as other proteins, is the description of their dynamical behavior at the molecular level. Unfortunately, an experimental technique that can provide a fully detailed molecular analysis of the time evolution of these proteins is not yet available and a combination of indirect evidences have to be used instead. Many experimental techniques are there and can provide different dynamic aspects. Depending on the lifetime of the event to be traced, different spectroscopic

(a) *KD in complex with substrate and AMP-PNP*

(b) *AMP-PNP binding, detail of adenine region*

(c) *AMP-PNP binding, detail of phosphates and Tyr substrate region*
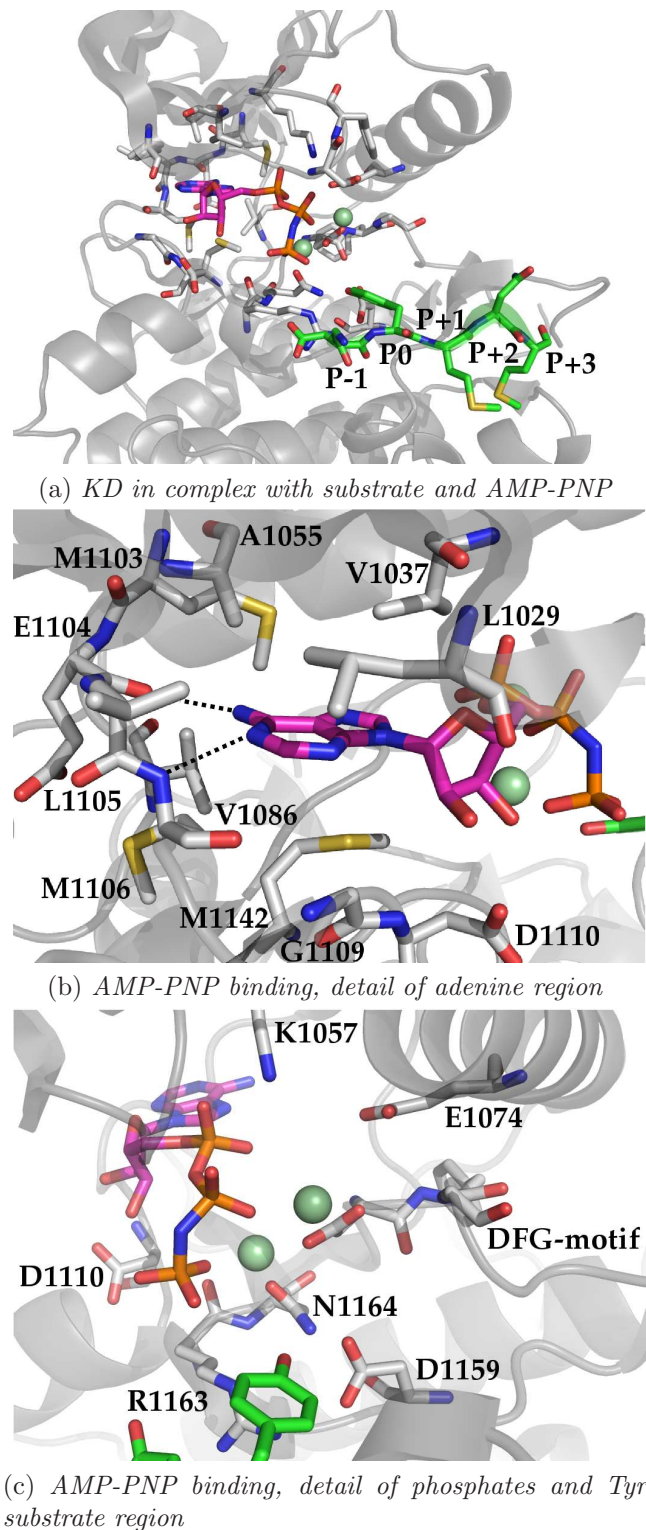
Figure 1.8: *Binding mode of AMP-PNP and substrate to insulin receptor. The secondary structure of the KD is represented as cartoon colored in gray. Atoms of ligands (AMP-PNP and peptidic substrate) as well as some residues of the ATP- and peptide-binding sites are in sticks representation and labeled. Magnesium ions are shown as spheres. The atoms are colored coded: blue (nitrogen), red (oxygen), yellow (sulfur), pale green (magnesium) and green, magenta, and gray (substrate, AMP-PNP and protein residues, respectively). Two hydrogen bonds of the adenine with the hinge region are depicted as dashed lines. For the substrate P0 is the position for the residue to be phosphorylated, in respect to this position P+n or P-n are residues C-terminal and N-terminal, respectively. Atomic coordinates (PDB code: 1IR3) rendered with the program* PYMOL *[64].*

methods can be used, for example CD, anisotropy resonance and mass spectrometry (for review [65]). The size of the protein and its variation can be followed by small-angle scattering. NMR is the only approach whose results are the explicit dynamics of proteins but so far it is still affected by the protein size problem and KD is beyond the limit of such investigation. Historically, dynamics of this domain have been deduced by comparison of different states of the protein solved by X-ray crystallography. This largely used method to investigate protein structure allows to look at the atomic disposition resulting in the Cartesian coordinates, but can be applied only when an energetic minimum is available in crystal form. Thus, it is clear that the "breath" of the proteins in solution cannot be considered.

The comparison of the active and inactive conformations reveals an enzyme activation accompanied by conformational changes of certain substructures (see figure 1.6 on page 14). The major movements are appreciated for three protein elements: C-helix, P-loop and A-loop, and the change of one toward the other two is clearly coupled [77]. The A-loop, undergoes structural change from a close conformation (figure 1.6a on page 14), when it interacts with the mouth of the protein, to an open conformation (figure 1.6b on page 14), folding on the side of the C-lobe. The variation of A-loop has two effects: one is to make the reaction center accessible to the reagents and the second is the formation of the platform for the substrate. At the N-terminus of the A-loop, the DFG motif switches from Asp-out to Asp-in (described in section 1.3.2 on page 8).
The nucleotide-binding loop has been described as a flap that changes conformation to gate and bind the cofactor at the reaction center. A taste of the dynamics of this element can be obtained by the comparison of crystallographic structures of Pka, reviewed by Taylor and coworkers [78].
The C-helix moves in order to bring its conserved aspartate (see section 1.3.1 on page 8) to the right distance for coordinating the catalytic lysine from the 3rd β-strand, which is then correctly oriented to bind the α and β-phosphates of the ATP. The conformational change of this element is defined by two limit conformations, the "swung-in" of the active state when is in the right position for catalysis (see figure 1.6a on page 14) and the "swung-out" conformation reached with a combination of movements and rotations, around its main axis, out of the protein core (see figure 1.6b on page 14) [69].

The rest of the secondary structure elements is moving in a concerted way with the hydrophobic core of each lobe. The general protein conformation is affected by subtle structural changes concerning the relative orientation and rotation of the two lobes. Two protein regions act as hinges, the loop between the C-helix and the 4th strand and the hinge segment. The movements of the two subdomains are independent for inactive forms, become coupled to a certain extent when the activation occurs and more when ligand at the ATP-binding site is present [65]. ATP is like a glue for the two lobes; in the active form, the majority of the interactions between the two lobes are mediated by the cofactor. These movements of the KD are likely to be important components of the protein kinetics. First, this structural breath allows a higher accessibility to the ATP-binding site favoring the cofactor intake. Second, being the distribution of the catalytic residues on both lobes, their positioning for reaction catalysis can be fine tuned. Last but not least, the presence of such structural vibrations are likely to affect the turnover number assisting

the expulsion of the products.

In-silico

Out of the wet-lab, the *in-silico* methodologies to investigate protein movements are several, for example there are stochastic approaches, like Montecarlo, and deterministic ones, such as molecular dynamics (MD) simulation. In general, the protein is described by atomic coordinates and charges, then several cycles of computation are run and the obtained trajectory represents possible states of the evolution of the system. These are highly fascinating techniques, but the price to pay for their applications is the reliability of the results, being aware of the several intrinsic limits, namely the poor treatment of atoms and bonds, and in the case of MD the disfavored ratio between calculated time and the time needed for the simulation.

Many reports of molecular dynamics regarding to the kinase domain can be found in the scientific literature. Here some of these studies are presented as well as their achievement on the dynamical behavior of KDs.
One of the first paper on the field described the MD simulation of cAMP-dependent kinase for 600 picoseconds (ps) [79]. The authors monitored the radius of gyration, flexibility, hydrogen-bond networking and the correlated motions of the protein during the calculated period of time. The results revealed that the large and small lobe acting as rigid bodies and conformational changes at the hinge region concern a group of amino acids instead of a single one.
In the year 2001, two studies with longer simulations were published. In the first, a homology model of cyclin-dependent kinase 1 (Cdc2 or Cdk1) was generated and parameters of active site atoms optimized for a 1 nanosecond (ns) long simulation [80]. The second is a 4 ns study of the dynamic coupling between the Src tyrosine kinase and its regulatory domains, SH2 and SH3 [81]. The authors, investigated the inhibition of the two regulatory domains hindering the motions of the KD of Src and Hck protein tyrosine kinases. They showed that SH2 and SH3 when in complex with KD act as a unity and suggested a main role in dynamics for the linker connecting these domains. The same dynamic behavior was assessed with an identical protocol in another non-receptor tyrosine kinase, Abl, carrying the 2 regulatory domains [82].
An atomistic view of the ADP released from Pka was proposed with pulling experiments during MD by Lu and coworkers [83]. With the applications of external forces to a certain cluster of atoms, the authors traced a possible way of ADP escape.
In a well described application of computational analysis, a comparison of Cdk2 thermal motions extracted from different sources is presented [84]. Ensembles of KD conformations were, generated by means of two computational approaches (one of which is MD) and taken from X-ray solved structures. The mapping of the inactive-active state transaction with virtual phosphorylation of a threonine in the A-loop was unsuccessfully attempted. In this paper the authors suggested that coulomb interactions can be overestimated in crystal structure in respect to MD simulations because of the crystallization conditions which might not fully correspond to the solution environment. In contrast, MD is supposed to better agree with the reality of this physical aspect because of the competitive thermal motions disruptions. To address the issue, simulations of virtual mutations of the three arginines stabilizing the negative charge of the phosphorylated threonine (in the

A-loop) were performed. The component interaction of one arginine is the major, while a second is described as the failsafe mechanism when the previous is weakened and the last residue seemed to make little contribution for the stabilization. This suggestion was partially confirmed with the degree of conservation of the three arginines through the kinome which might be linked with the importance of the residue function [84].

Another study on the transition between inactive and active conformations and the reverse reaction was carried out with a special flavor of MD [85]. Here, Lyn tyrosine kinase has been investigated with Biased Molecular Dynamics where a biased potential is applied to the molecular mechanics calculation to force part of the protein to predestined pathway. The outcome demonstrated that in the KD a special network of electrostatic interactions switches to an alternative pattern undergoing activation and this is proposed as the mechanism driving the conformational change. The six residues involved in this networking[7] are shown to be well conserved in PKs and part of secondary structure elements supposed to play major roles in KD activation: 3rd helix, A-loop, C-helix and catalytic loop [85].

1.3.4   The catalysis

The putative reaction catalyzed by these enzymes is depicted in figure 1.9 on the facing page. This picture of the phosphotransfer does not come from any time-resolved experiment, instead it has to be considered more like a tool for helping the reader's understanding. Although, all hydrogen atoms are missing and the view is restricted to few reaction players the picture describes the primary function of the kinase domain. From this picture it becomes clear why the right positioning and orientation of the two reagents is so important. The proximity of these two entities permits the breakage of the bond between the oxygen of the β-phosphate and the phosphorus of the γ one (figure 1.9a on the next page), the orbital configuration inversion of the latter (in figure 1.9b) and eventually the formation of the new bond with the oxygen of the hydroxyl group of the substrate (in figure 1.9c). The transfer turnover was extrapolated to be equal to 500 sec^{-1} for Pka underlining the efficacy of this catalytic system [87].

Despite members of the PK superfamily catalyze the same reaction, a broad variety of kinetics differences exists and characterizes different kinases. When kinetics experiments are run with the Pka in solution with ATP, small peptides and Magnesium, a cycle of the reaction for this protein is yielded in millisecond timescale (turnover number = 45 sec^{-1} [87]). The difference between the transfer rate and the catalytic turnover can be attributed to the "environment effect", in other words, influenced by the surrounding of the reaction.

A catalytic cycle can be ideally broken into steps as follow:

- the protein binds ATP with its coordinated Magnesium ions and the peptide substrate almost randomly [88];

- then the transfer of the γ-phosphate of the cofactor to the substrate occurs [87];

---

[7]Namely lysine of subdomain II, glutamate of subdomain III, aspartates of subdomains VII and VIII an arginine and a tyrosine (auto-phosphorylation site) of the A-loop.

(a) *reagents state*

(b) *intermediate state*

(c) *products state*
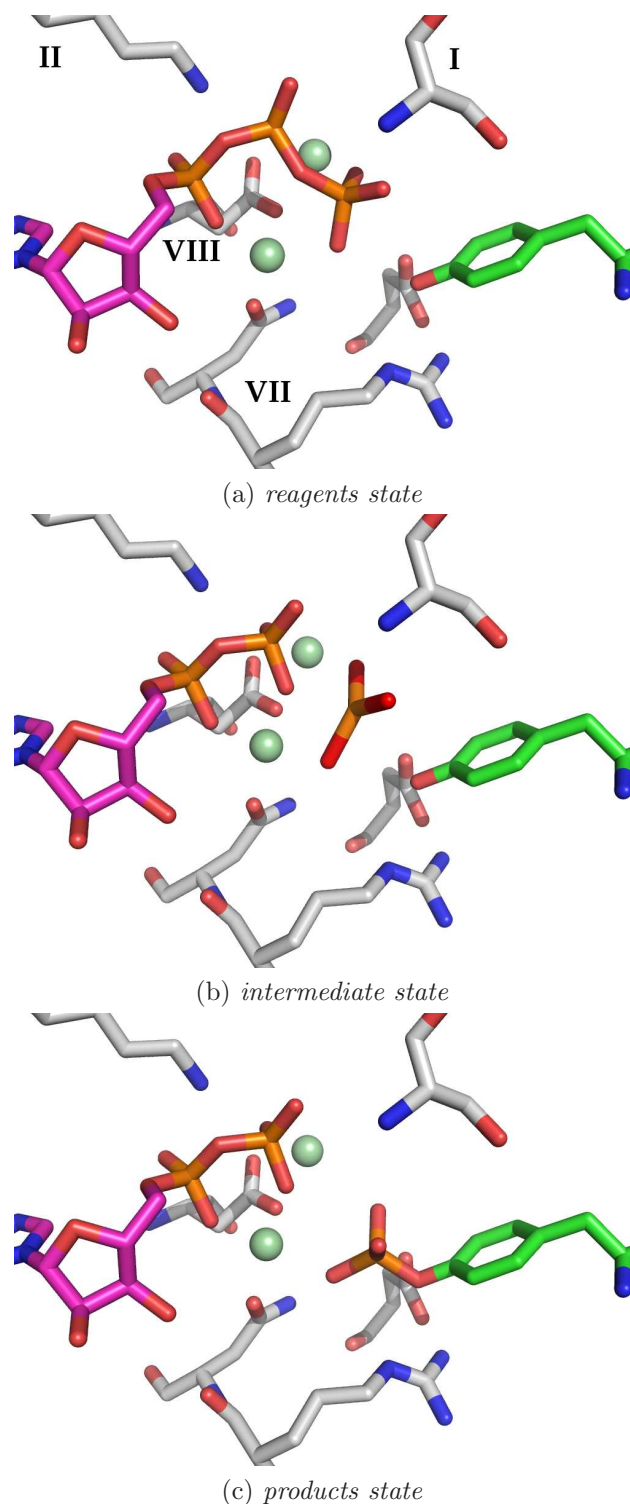
Figure 1.9: *Phosphotransfer reaction towards hydroxyl group of tyrosine residue at the substrate. The hypothesis of the chemical reaction is drawn in three steps, the scheme is adapted from [86] and hydrogen atoms are omitted (a, b, c). Atoms of reagents are in sticks representation as well as some protein residues (subdomains of provenance are labelled); Magnesium ions are shown as spheres. The atoms are colored coded: blue (nitrogen), red (oxygen), yellow (sulfur), pale green (magnesium) and green, magenta, and gray for carbon atoms of substrate, ATP and protein residues, respectively. Atomic coordinates are rendered with the program pymol [64].*

- the final step is the products release where the phosphopeptide leaves first [89] followed by the ADP.

Now the protein is free to start all over again. Thus, different aspects can influence the speed of catalysis as demonstrated by several biochemical studies.

Running kinetic experiments in different viscosity conditions, the rate limiting step of the reaction has been shown to be the ADP release for Pka [90, 91] and for insulin receptor [92] with activated proteins. Furthermore, Shaffer and Adams proposed two additional steps for Pka catalysis: a prior key residues positioning and their relaxation subsequent the transfer [89, 93].

Clearly, the protein conformational changes can regulate the kinetics of the phosphoryl-transfer by reaction tuning and products release.

The importance of Magnesium for the reaction was also tested and as expected from the chemistry of the protein reaction center, no kinetics could be recorded in absence of the ion assaying insulin receptor [94]. In Pka various concentrations of the Mg^{++} changed significantly the steady-state kinetic parameters, through the enhancement of ATP binding and affecting protein conformational changes [93]. However, these behaviors are representative for *in-vitro* conditions using peptides as substrates, and a different panorama is achieved when native substrates are used instead of small peptides [95].

Another interesting aspect about the phosphoryl-transfer that has been investigated, is the transition state of the reaction. When the phosphoryl-donor and acceptor are at the right distance, the oxygen atom of the substrate hydroxyl group (the nucleophile) attacks the phosphorus of the ATP γ-phosphate with a partial bond formation between these two entities. Then, the transition state occurs with the weakening of the covalent bond between the second and the third phosphates of ATP and the concomitant placement of the latter one toward the nucleophile. Depending on the bonds character and the distance between the species, two transition states are possible as depicted in figure 1.10 on the facing page [96, 97].

Dissociative transition state: which is characterized by a little bond formation between the phosphorus and the hydroxyl group, and a largely broken bond between the latter and the β-phosphate of the ADP.

Associative transition state: in this alternative state, the character of the two bonds is stronger with reactants closer to each other.

Secondary differences between the two transition states are i) the required nucleophilicity of the attacking oxygen atom and ii) the role played by the aspartate of the C-loop.

Experimentally, in agreement with the associative reaction, no dependence was found when Pka activity was measured in an pH interval from 6 to 9 [98]. The authors considering the pKa values for aspartate and serine suggested that the protonation state of the aspartate might not be essential for the reaction.

The dissociative transition state of the reaction have been hypothesized studying linear-free energy relationships in determination of the nucleophile coefficient for the reaction [96]. The authors studied the pKa and kinetics at different pH using fluorinated tyrosine analogues [96].
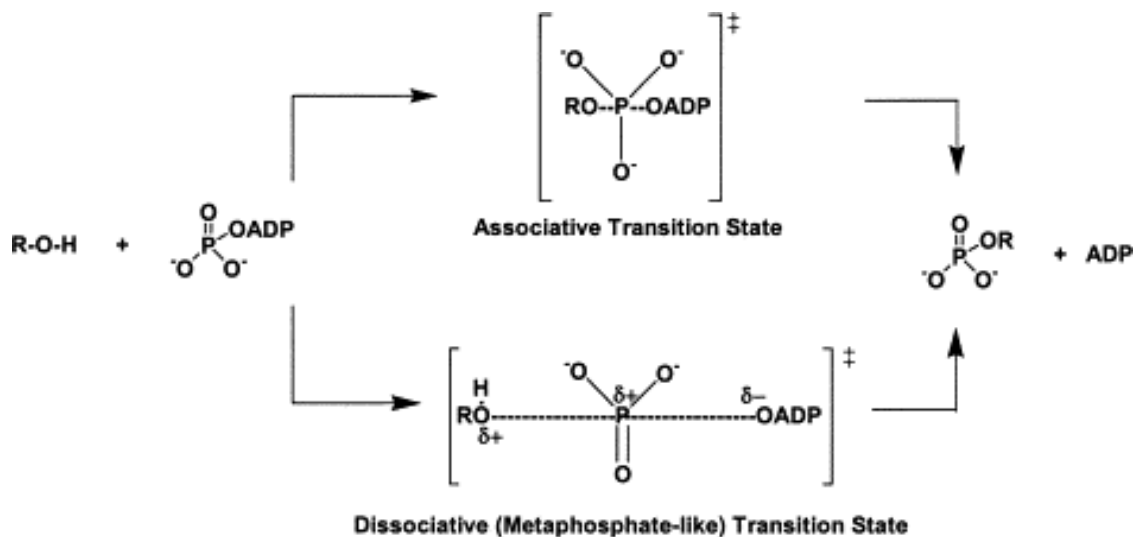
Figure 1.10: *Transition states of the phosphoryl-transfer reaction. Associative and Dissociative transition states are depicted in two-dimensional scheme, taken from [97].*

In-silico

The fascinating atomistic point of view of the reaction can be hypothesized by experiment outcomes but can also be investigated theoretically. The chemistry of the phosphoryl-transfer reaction have been studied computationally by means of several methods with different level of theory. However, whether the key atoms for reaction are simulated with explicit or implicit electron models, and whether or not their atomic surroundings are included, can have a big impact on the outcomes.

Early calculations by means of semi-empirical methods proposed a high activation energy barrier for the reaction [99–101]. But later, the inadequate use of these methods for studying the kinase reaction was shown [102]. The quanto-mechanics (QM) studies of the reaction catalyzed by Pka [103–107] and Cdk2 [108] debated about the "associative" and the "dissociative" transition state with the involvement of the conserved aspartate of subdomain VII. A minor role for this residue in positioning the phosphoryl-acceptor have been suggested as well as the primary one where it takes part at the reaction as catalytic base deprotonating the hydroxyl group and enhancing its nucleophilicity. Higher description accuracy of the events comes from recent applications of the mixed type quanto- and molecular-mechanics (QM/MM) calculations [109, 110] where atoms directly involved in the reaction and those in the surroundings are described with QM and MM models, respectively. In agreement with the dissociative transition state and with a major role played by the aspartate, Chen et al., [109] calculated a free energy barrier for the Pka reaction around 14 kcal/mol. While, the later study on Cdk2 confirms the associative fashion for the reaction with the minor role for the aspartate and an activation energy of approximately 24 kcal/mol [110].

1.3.5   The auto-regulative inhibition

For the crucial role of protein kinases in the cell physiology, many different types of controls of their catalytic domain have evolved. The negative regulation of KD activity is expressed by a large variety of inhibition mechanisms and here few examples are reported.

- Cdks are inactive protein serine kinases (PSKs) in the basal state but when interacting with cyclins through the C-helix they became active [111].

- Pka kinase is inhibited by an endogenous peptide, called protein kinase inhibitor (PKI), which binds and traps the KD at the substrate site as depicted in figure 1.11a on the next page [45, 112].

- The Ca^{++}-regulated kinases Twitchin possesses a C-terminal tail substructure which folds back into the ATP and the substrate-binding sites, blocking both [113] (figure 1.11b on the facing page).

- Src tyrosine kinase comprises beside the KD two regulatory domains at the N-terminal, called SH2 and SH3, which interact with the two lobes of the KD. This, limits the freedom of the lobe's movements and maintain, via internal interactions, the C-helix in a swung-out conformation [114–116] (figure 1.11c).

- Flt3, a receptor tyrosine kinase, possesses a domain N-terminal to the KD, named juxtamembrane domain, which interacts with the C-helix and then folds in the interface between the two lobes. In this way the A-loop cannot adopt the open conformation [117] (figure 1.11d and further discussed in details).

1.4   Tyrosine Kinases

Growth, differentiation and intercommunication of cells are associated with tyrosine phosphorylation. Stable machineries, PTKs, to trigger it have been discovered already in the early stages of multicellularity for eukaryotic organisms (see subsection 1.1.4). The primary role played by protein tyrosine kinases in cell physiology was immediately understood when elements of the family were discovered in tumor virus [47–49], and their presences associated with overgrowth and morphological changes of cells occurring when infected by one of these virus, Rous sarcoma virus [118–120]. Later, the hypothesis was supported by the finding of protein tyrosine kinase activity for the epidermal growth factor receptor [121], insulin receptor [122] and platelet-derived growth factor receptor [123]. Another interesting aspect of TK signaling concerns the substrates. Peptides containing phosphotyrosine residues were discovered as ligands for specific domains as SH2 and PTB that are involved in cell signaling pathways [55, 124].

Today, two classes of tyrosine kinases are known: the receptor tyrosine kinases (RTKs), mainly single polypeptide spanning the cellular membrane, and the non-receptor tyrosine kinases (nRTKs), differently localized inside the cell. A large portion of scientific literature has been devoted to study structure, function, interaction as well as other behaviors

(a) *PKA + PKI (1ATP)*

(b) *Twitchin kinase (1KOB)*

(c) *Src kinase (1KSW)*

(d) *Flt3 kinase (1RJB)*
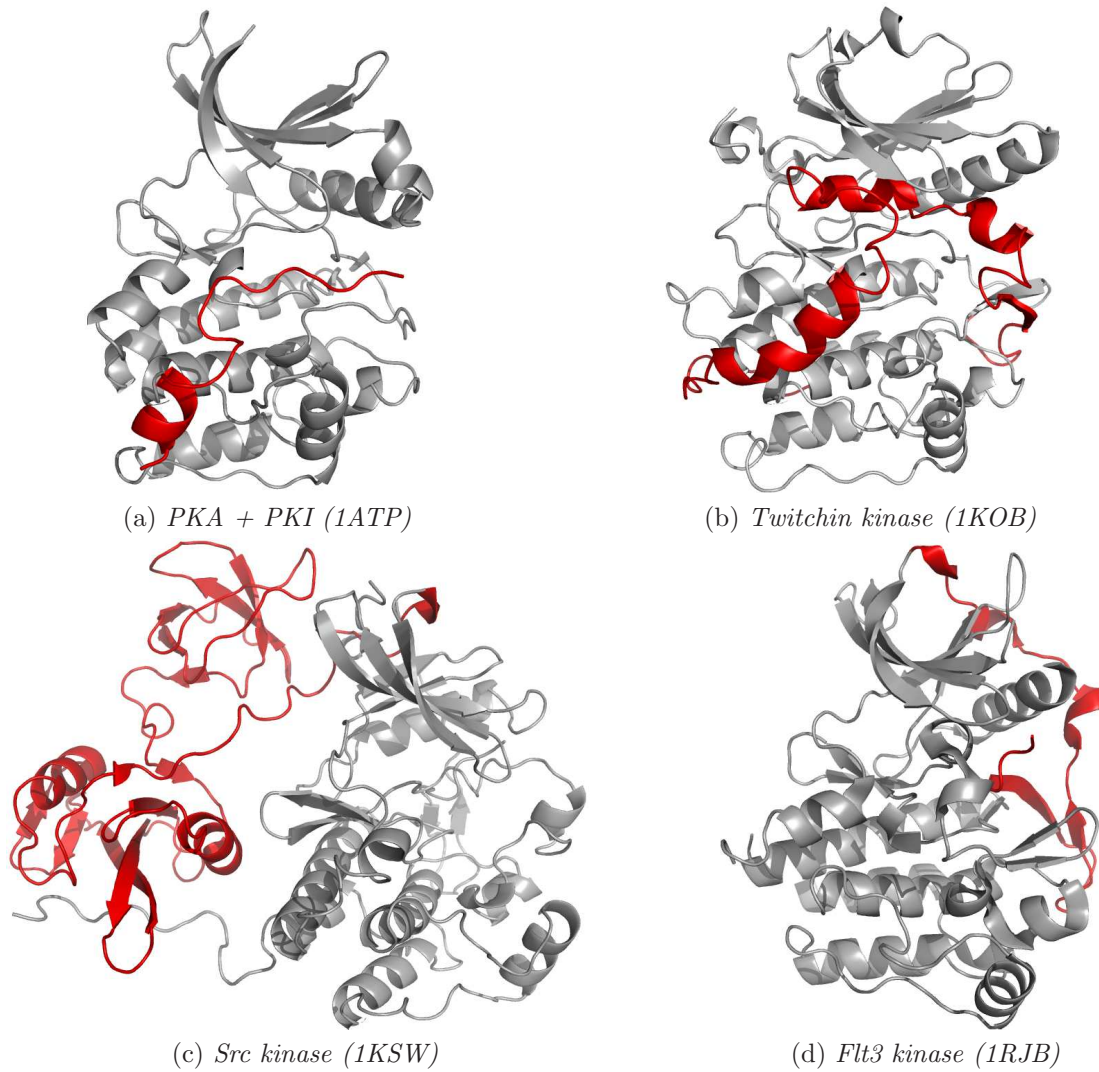
Figure 1.11: *Inhibitions mechanisms for two PSKs (a and b) and two PTKs (c and d). The secondary structure of the four kinase domains is represented as gray cartoon. The secondary structure of protein kinase inhibitor (PKI), C-terminal tail of Twitchin kinase, SH2 and SH3 domains of Src and the juxtamembrane domain of Flt3 is represented as red cartoon. Atomic coordinates of PDB files (reported in between brackets) are rendered with the program* PYMOL *[64].*

of PTKs. Their component in cell-signaling has been deeply investigated. Examples of their regulatory activity for intra- and intercellular stimuli are response to insulin, the vascularization process, the differentiation of the immune system, the development of the central nervous system and many many others [17, 36, 37].

The central role assigned to these proteins in cell physiology suggests the importance of the tight regulation of their activity. Although, protein tyrosine phosphatases are counteracting kinase activity, series of inhibition mechanisms are intrinsically implemented to maintain low productivity for these enzymes. The loss of control of the tyrosine kinase activity leads usually to uncontrolled growth and differentiation and thus to malignant transformation of cells [39]. Uncontrolled kinase activity can be the result of mutations or deletions of one or more residues, protein overexpression and chromosomal translocation leading to the fusion of part of the kinase polypeptide with other proteins [39, 125].
In some cases these types of alterations are causatively linked to oncogenesis or cancer progression [126].

The 90 human sequences assigned as protein tyrosine kinases include 58 RTKs divided in 20 subfamilies and 32 nRTKs divided in 10 subfamilies [38, 127].

1.4.1   Receptor tyrosine kinases

The 20 subfamilies composing the receptor tyrosine kinase superfamily are listed in table 1.1.

Table 1.1: *The 20 RTKs subfamilies and their members*

Subfamily	Members	Entry code
PDGFR	**Pdgfrα**: Platelet-derived growth factor receptor α	P16234
	Pdgfrβ: Platelet-derived growth factor receptor β	P09619
	Cdf1r: Colony-stimulation factor-1 receptor	P07333
	Kit: Stem-cell factor receptor	P10721
	Flt3: Fms-like kinase receptor	P36888
EGFR	**EGFR1**: Epidermal growth factor receptor 1	P00533
	ERBB2: Epidermal growth factor receptor 2	P04626
	ERBB3: Epidermal growth factor receptor 3	P21860
	ERBB4: Epidermal growth factor receptor 4	Q15303
IR	**INSR**:Insulin receptor	P06213
	IGF1R: Insulin-like growth factor 1 receptor	P08069
	INSRR: Insulin receptor-related protein	P14616
VEGFR	**Vegfr1**: Vascular endothelial growth factor receptor 1	P17948
	Vegfr2: Vascular endothelial growth factor receptor 2	P35968
	Vegfr3: Vascular endothelial growth factor receptor 3	P35916
FGFR	**Fgfr1**: Fibroblast growth factor receptor 1	P11362
	Fgfr2: Fibroblast growth factor receptor 2	P21802

Subfamily	Members	Entry code
	Fgfr3: Fibroblast growth factor receptor 3	P22607
EPHR	**EphrA1**: Ephrin receptors A1	P21709
	EphrA2: Ephrin receptors A2	P29317
	EphrA3: Ephrin receptors A3	P29320
	EphrA4: Ephrin receptors A4	P54764
	EphrA5: Ephrin receptors A5	P54756
	EphrA6: Ephrin receptors A6	Q15375
	EphrA7: Ephrin receptors A7	P29322
	EphrA8: Ephrin receptors A8	Q5JZY3
	EphrB1: Ephrin receptors B1	P54762
	EphrB2: Ephrin receptors B2	P29323
	EphrB3: Ephrin receptors B3	P54753
	EphrB4: Ephrin receptors B4	P54760
	EphrB5: Ephrin receptors B5	O15197
CCK	**Ptk7**: Tyrosine-protein kinase-like 7 Cck4	Q13308
NGFR	**Ntrk1**: Neurotrophic tyrosine kinase receptor type 1	P04629
	Ntrk2: Neurotrophic tyrosine kinase receptor type 2	Q16620
	Ntrk3: Neurotrophic tyrosine kinase receptor type 3	Q16288
HGFR	**Met**: Hepatocyte growth factor receptor	P08581
	Ron: Macrophage-stimulating protein receptor	Q04912
AXL	**Ufo**: Tyrosine-protein kinase receptor UFO	P30530
	Merk: Proto-oncogene tyrosine-protein kinase MER	Q12866
	Tyro3: Tyrosine-protein kinase receptor TYRO3	Q06418
TIE	**Tie**: Tyrosine-protein kinase receptor Tie-1	P35590
	Tie2: Tunica interna endothelial cell kinase Tek	Q02763
RYK	**Ryk**: Tyrosine-protein kinase RYK	P34925
DDR	**Ddr1**: Epithelial discoidin domain-containing receptor 1	Q08345
	Ddr2: Epithelial discoidin domain-containing receptor 2	Q16832
RET	**Ret**: Proto-oncogene tyrosine-protein kinase receptor ret	P07949
ROS	**Ros**: Proto-oncogene tyrosine-protein kinase ROS	P08922
LTK	**Ltk**: Leukocyte tyrosine kinase receptor	P29376
	Alk: Anaplastic lymphoma kinase	Q9UM73
ROR	**Ror1**: Tyrosine-protein kinase transmembrane receptor ROR1	Q01973
	Ror2: Tyrosine-protein kinase transmembrane receptor ROR2	Q01974
MUSK	**Musk**: Muscle, skeletal receptor tyrosine protein kinase	O15146
LMR	**LMTK1**: Lemur tyrosine kinase 1	Q6ZMQ8

Subfamily	Members	Entry code
	LMTK2: Lemur tyrosine kinase 2	Q8IWU2
	LMTK3: Lemur tyrosine kinase 3	Q96Q04
SuRTK106	**STYK1**: Tyrosine protein-kinase STYK1	Q6J9G0

Many extracellular stimuli, like growth factors, differentiation factors and hormones, important for the regulation of the normal development of multicellular organisms, impose their actions via binding the specific receptor tyrosine kinases at the cell surface [128]. In general, this interaction event is correlated with the non-covalent dimerization of receptor units. The signal is transduced inside the cell by initiating the activity of the kinase domain which is linked, via transmembrane domain, to the external ligand-binding domain. Eventually, the kinase domain phosphorylates tyrosine residues in the receptor itself or in secondary proteins.

The ligand-binding domain is usually composed by different combinations of domain sequences, like immunoglobulin-like domain, cysteine-rich region, fibronectin type III domain and others [130] and differs between the RTK families. The transmembrane domain (TM) is similar for all RTKs and assumes an α-helical fold in the phospholipidic bilayer. A stretch of amino acids, called juxtamembrane domain (JM), connects TM to the catalytic domain. The JM of several RTKs has auto-inhibitory function, mainly hindering the C-helix movement [73, 117, 131]. The stable kinase activity is achieved with the phosphorylation of conserved tyrosine residues located on this domain. This mechanism was confirmed by the catalytically repressed protein form obtained with mutations of these residues in Ephr [132], Pdgfr [133, 134] and Musk [135]. The last receptor portion, the C-terminal tail, is conserved for each family especially for the positions of certain tyrosine residues [136], in some cases they have been shown to be auto-phosphorylation sites [137]. From crystallographic evidence the C-terminal tail of Tie2 RTK can act as auto-inhibitory mechanism [138] which is confirmed by the elevated kinase activity of the C-terminal deletion [139].

For the ligands interacting with the extracellular binding region of the protein, several types of polypeptide chain and ways of interacting have been described. The platelet-derived growth factor (Pdgf) and the vascular endothelial growth factor (Vegf) are homodimeric proteins linked via disulfide bridge and binding two different receptor monomers at the same time [130]. The epidermal growth factor interacts non-covalently with one RTK monomer producing quaternary structure rearrangements and subsequent binding to a second monomer [140, 141]. For inducing receptor dimerization, the fibroblast growth factor requires the binding of a second molecule, the heparin sulfate proteoglycan [142, 143]. Accessory proteins supporting receptor dimerization upon ligand-binding have been revealed even in the case of Ret and Musk [144, 145]. In the case of insulin the mechanism is different. The insulin receptor "dimer" is always present and the two units are formed by α and β chains interlinked by a disulfide bridge. The α chain contains the extracellular domain while the β one the transmembrane, juxtamembrane and catalytic domain. Thus, this receptor is found at the cell membrane as heterotetramers, $\alpha_2\beta_2$, in normal conditions and the insulin binding causes a quaternary structure rearrangement of the tetramer, followed by receptor's activation.
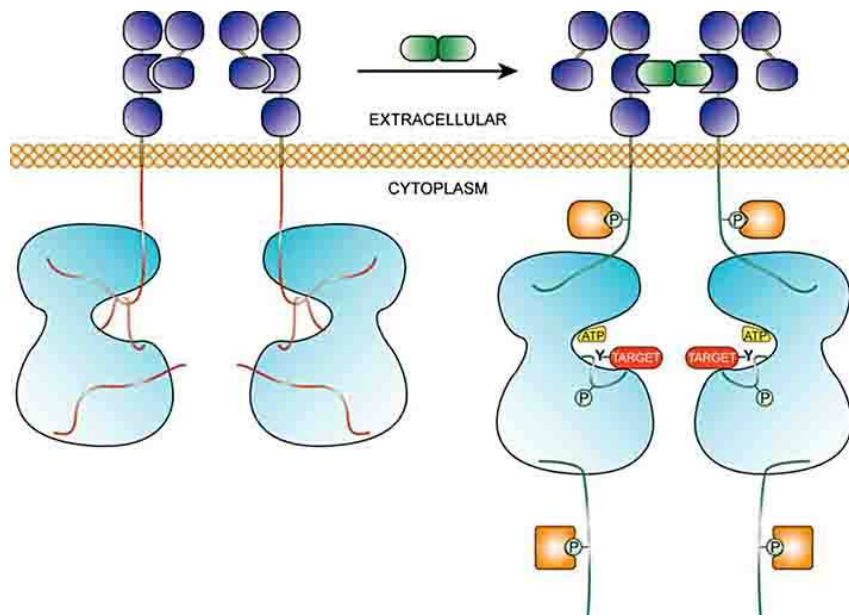
Figure 1.12: *Activation of RTKs. Various protein portions and ligands are depicted and colored, kinase domain in cyan, extracellular domains in purple, ligand in green, cofactor in yellow, substrate in red, downstream signaling proteins in orange. Other cytoplasmatic protein portions are drawn, namely juxtamembrane, A-loop and C-terminal tail. The ligand-free and the activated states are described in the left and right side, respectively. Picture taken from [129].*

The extracellular part of these receptors characterizes the RTKs families and captures the ligand selecting the stimulus to be entered in the transducing machinery. The signal crosses the cell membrane by the formation of the ligand-assisted receptor dimerization, which increases the local concentration of the cytoplasmic domains favoring the trans-autophosphorylation of tyrosine residues located at the A-loop of KD, juxtamembrane domain and C-terminal tail [146, 147] (depicted in figure 1.12). The phosphorylated receptor is now able to transfer the signal in the cytoplasmatic environment releasing it to secondary effectors such as other protein kinases or phospho-tyrosine binding (PTB) domains.

The signal transfer is a combination of two mechanisms. The first one includes the partner recruitment via phosphorylated tyrosine residues in the catalytic domain, juxtamembrane domain and the C-terminal tail. The second one is related to the phosphorylation of specific sequences based on the substrate selection of the KD as discussed earlier (see section 1.3.2 on page 8). For example, proteins containing SH2 and PTB domains can be selected and activated with conformational changes due to the interaction or phosphorylation of the KD, being localized in close proximity [39, 147].

Studies on chimera receptors were performed to understand how the extracellular ligand-binding and the intracellular catalytic domain are influencing one each other. Chimera receptors mixing protein extra- and intracellular portions of different kinases were expressed and tested. These hybrid receptors could bind one type of ligand according to the ligand-binding domain and arise different stimuli according to the cytoplasmatic domains [148–150]. Thus, the results revealed that incoming stimuli and out-coming signals

of RTKs are rather independent mechanisms and that the intracellular part is decisive for the type of signal that will be forwarded within the cell.

The signal flux is readily interrupted with the endocytosis of the receptor-ligand complex. At this stage the proteins inside the vesicles can be degraded by lysozime or be recycled at the cell surface [128].

Although the general fold of the kinase domain is conserved, a remarkably large amino acid insertion, called kinase insert region, is observed in the loop connecting the first two helices ($\alpha2$ and $\alpha3$) of the C-lobe (figure 1.5 on page 13) of some receptors tyrosine kinase, namely the Pdgfr-family and Vegfr-family elements. This sequence has high variability in length and residues composition. The length can vary from 10 to 100 residues among TKs [128, 130]. It appears more conserved for orthologue proteins, same protein type from different species, than for homologue members of the same subfamily [128]. No secondary structure elements have been so far depicted by any experiments. Furthermore, the deletion of part of this loop in Vegfr1 permitted the protein crystallization and its structure solution via X-ray [151], resulting in the common KD architecture. This region have been described essential for arising certain TK pathways [152], but the function is still not fully established [153].

Another peculiarity of RTKs is that some of them required A-loop phosphorylation for full kinase activity [154–156], while others, like Ret and Egfr do not [157, 158]. In the case of Egf receptor, catalytic domain shows a Cdk behavior, requiring protein-protein interaction involving the C-helix for activation [72]. While cyclins fulfill this task by the Cdks, a second KD of Egfr seems to interact similarly to cyclins with the first one [72]. To confirm this observation, it could be shown that an intrinsically inactive KD of Egfr3 is able to activate other family members forming heterodimers [159].

There are several examples of pathologies involving disfunctions of RTKs, The Egfr over-expression has a causative role in human glioblastoma, mammary carcinomas and lung cancer [39, 160]. In addition, Egfr sequence mutations are found in lung cancer patients [161]. Different mutations localized at extracellular and intracellular portions of Kit are correlated with the development of gastrointestinal stromal tumors (GIST) [162]. The deregulation of Flt3 activity is found in 30% of cases of acute myelogenus leukemias (AML) due to additive sequences, called internal tandem duplication (ITD) [163]. Gene translocation expressing the fusion of N-terminal of Nucleophosmin (Npm) and KD of Alk, so-called Npm-Alk, is associated with the development of anaplastic large cell lymphoma (ALCL) [164].

Furthermore, RTKs have also been described in genesis or progression of other diseases, like diabetic retinopaty, atherosclerosis and psoriasis [165].

1.4.2 non-Receptor tyrosine kinases

The second group of protein tyrosine kinases is represented by cytosolic proteins, though someone (e.g. Abl) can have dual distribution with their presence also in the nuclear environment [17]. The 10 subfamilies composing the non-receptor tyrosine kinase superfamily are listed in table 1.2 on the next page.

Table 1.2: *The 10 nRTKs subfamilies and their members*

Subfamily	Members	Entry code
SYK	**Ksyk**: Tyrosine-protein kinase SYK	P43405
	Zap70 70 kDa zeta-associated protein kinase	P43403
FRK	**Frk**: FYN-related kinase	P42685
	Ptk6: Breast tumor kinase	Q13882
	Srms: Tyrosine-protein kinase Srms	Q9H3Y6
FES	**Fes**: Proto-oncogene tyrosine-protein kinase Fes/Fps	P07332
	Fer: Proto-oncogene tyrosine-protein kinase FER	P16591
FAK	**Fak1**: Focal adhesion kinase 1	Q05397
	Fak2: Focal adhesion kinase 2	Q14289
CSK	**Csk**: C-terminal Src kinase	P41240
	Matk: Megakaryocyte-associated tyrosine-protein kinase	P42679
ACK	**Ack**: Activated CDC42 kinase 1	Q07912
	Tnk1: Non-receptor tyrosine-protein kinase TNK1	Q13470
JAK	**Jak1**: Janus kinase 1	P23458
	Jak2: Janus kinase 2	O60674
	Jak3: Janus kinase 3	P52333
	Tyk2: Non-receptor tyrosine-protein kinase TYK2	P29597
TEC	**Tec**: Tyrosine-protein kinase Tec	P42680
	Bmx: Bone marrow tyrosine kinase gene in chromosome X protein	P51813
	Btk: Bruton tyrosine kinase	Q06187
	Itk: Tyrosine-protein kinase ITK/TSK	Q08881
	Txk: Tyrosine-protein kinase TXK	P42681
ABL	**Abl1**: Abelson murine leukemia viral oncogene homolog 1	P00519
	Abl2: Abelson-related gene protein	P42684
SRC	**Src**: Proto-oncogene tyrosine-protein kinase Src	P12931
	Lck: Lymphocyte cell-specific protein-tyrosine kinase	P06239
	Fgr: Proto-oncogene tyrosine-protein kinase FGR	P09769
	Fyn: Proto-oncogene tyrosine-protein kinase Fyn	P06241
	Yes1: Proto-oncogene tyrosine-protein kinase Yes	P07947
	Blk: B lymphocyte kinase	P51451
	Hcks: Hemopoietic cell kinase	P08631
	Lyn: Tyrosine-protein kinase Lyn	P07948

These TKs participate in cell signaling as well as the RTKs [22, 166]. Within the signaling pathways, they have a dual role. On the one hand, they act as second effectors of intra- and intercellular stimuli regulating interactions and activity of other proteins, like kinases (as component of the kinase cascades). On the other hand, some nRTKs, like

Src, Fak and Jaks, associate with transmembrane proteins such as cytokines and integrin adhesion receptors fostering synergistically the signal transduction of membrane bound proteins [167–169]. The activation mechanism of these binary complex receptors resembles the one previously described for RTKs, where protein oligomerization stabilizes KD proximity for tyrosine phosphorylation [167, 168].

The overall architecture of these proteins is characterized by the presence of different domains modules at the N- and C-terminal position of the catalytic domain. The most frequently found domains are the SH2, the SH3, the Actin-binding, the DNA-binding, the Focal adhesin-binding, the Integrin-binding, the Pleckstrin homology (PH) and the Btk domains. Different combination and distribution of such protein portions diversified the topology of the elements of the enzyme family. These ubiquitous modules act through a variety of concerted functions that are characteristic of them. Their main roles in nRTKs are regulation of the kinase activity, protein localization and interactions with secondary entities [146, 169, 170].

One of the best characterized nRTK, Src, is composed by SH3, SH2 and catalytic domain, and its behavior is reported here as example for illustrating the modular function of the different domains. The Src N-terminal interacts via myristoyl group to the inner surface of the cellular membrane where the protein is thus localized. The inactive state is maintained by the complex of the rest of the molecule. SH3, SH2 and KD domains interact with each other, depicted in the crystallographic structures [115], forming two SH2-KD, one SH3-KD and one SH2-SH3 interacting interfaces. The most prominent interaction is the one between the phosphotyrosine residue at the C-terminal tail of the KD and the SH2 domain. This residue is phosphorylated by another nRTK, Csk whose physiological role is indeed to keep Src inactive. With the protein activation, SH2 and SH3 are released and can interact with other sites, while the KD achieves catalytic activity.
Different regulatory mechanism is promoted by SH2 and SH3 domains in the case of Csk where their presence is required for full kinase activity [171]. The crystal structures (PDB codes: 1BYG and 1K9A) comparison shows that: the regulatory domains interact with the N-terminal lobe of the catalytic domain [172] stabilizing the catalytic favorable conformation of the C-helix [173].

Two examples of modified nRTKs associated with diseases are reported here. The constitutive activation of Src due to the aberrant deletion of its C-terminal tail has been correlated in some colon cancer cases, while Src over-expression has been depicted in pancreatic cancers [39]. The Philadelphia chromosome resulting from gene translocation leads to the expression of the fusion protein Bcr-Abl with uncontrolled tyrosine kinase activity, that has been found to be the causative agent of chronic myelogenus leukemia (CML) [126].

1.5 TK Inhibitors

The large implication of protein tyrosine kinases in many types of cancer have pushed scientists, from both academy and industry, to search for external control of these enzymes. The direct pharmacological targeting of tyrosine kinases in treatment of cancer is quite

recent and dates from the end of 1990s. Till late 1980s the cancer therapies consisted exclusively in cycles of physical and chemical applications localized or not. Radiotherapy, alkylating agents (nitrogen mustards, cisplatinum, etc.), microtubule destabilizing compounds (taxol, docetaxel, etc.), intercalating molecules (doxorubicin, methylene blue, etc.), topoisomerase inhibitors (topotecan, amsacrine, etc.) are therapies whose selectivity is solely based on the difference in cell growth and replication between normal and tumor cells. The consequence of the lack of selectivity is the intrinsic high toxicity and severe side effects for the patients.

A lot can be written regarding the many ways attempted to inhibit the tyrosine kinase activity, but here, the most explored ones are described.
One of the successful treatment against tyrosine kinases disregulation is the use of monoclonal antibodies against the extracellular domain [174]. The specific binding of the antibody hindered the interaction with the ligand and other receptor monomers [175], and thus, the receptor activation is blocked because of the abolished dimerization. Herceptin, an antibody against HER2 (Egfr), was accepted by Food & Drug Administration (FDA) in 1998 and is used for the treatment of breast cancer [176].

A second promising approach is the usage of the so-called mechanism-based inhibitors. These molecules interact at reaction center occupying the substrate and cofactor-binding sites. These bi-substrate inhibitors are the result of the chemical fusion of a peptide substrate and an ATP analogue with a good protein affinity, because of the summation of the interactions, and selectivity for the peptide mimetic [97]. The confirmation of their action at the molecular level was gained with the crystallographic structure of insulin receptor bound to one of such molecules [177]. Many attempts have been made and are reviewed in [97], but unfortunately none of these compounds made it into the clinic up to now.

The most explored and clinically successful TKs inhibition is based on adenine mimetic behavior of molecules competing with ATP for binding at the same protein pocket.
Several scaffolds have been exploited and among the most promising there are quinazolines, like Iressa and PD153035, phenylamino-pyrimidines, like Imatinib, pyridopyrimidines, like PD173074, pyrrolopirimidines, pyrazolopirimidine, like PP1 and PP2 and staurosporines like molecules [178]. Their main interaction moiety appears to be the "ligand-core" [75] with the conserved region of the ATP-binding site while specificity is achieved with the additional contacts in the less conserved surroundings, like the hydrophobic pocket [74, 179, 180].
The breakthrough in tyrosine kinase inhibition was the discovery of Imatinib, which shows selectivity against the oncogenic fusion protein Bcr-Abl, Kit and Pdgfr. Coming from a lead-identification screening against protein kinase C, it was later developed for better pharmacokinetic and pharmacodynamic properties [181]. During optimization the compound showed selective inhibition of Bcr-Abl fusion protein and clinical trials were performed as cure for Chronic Mylogenus Leukemia (CML), finally its approval for marketing by FDA came in 2001 [181].
Although the *in-vivo* efficacy of these inhibitors could suffer from the high concentration of ATP that is present in cells (8 mM) and even more in transformed cells [182], the clinical data on Imatinib clearly indicate that effective *in-vivo* inhibition can be achieved.

Thus, the positive results suggest that the issue of the targeted conformation is relevant for the *in-vivo* outcome.

Due to the highly conserved ATP-binding site through the protein kinase family [74], the selectivity of these inhibitors is a second major issue. Hitting more than one tyrosine kinases with the same molecule might be undesirable knowing the central physiological roles of these proteins. On the other hand, there are evidences for tyrosine kinase inhibitors which owe their therapeutic effectiveness to the PKs additional inhibited [183].

Cancer treatments based on the administration of ATP-competitive inhibitors are successful, as shown with Imatinib [181], but relapses can occur [184,185] due to alterations of the target. Resistant protein mutants to the inhibitor effect give to some cancer cells survival advantages [186]. Thus, it is a general opinion in the field that therapeutic strategies based on multiple drugs as it is the case for the AIDS treatment, will be needed in the future to cure the cancer. The design of small molecules for targeted cancer therapy is still the major trend in the scientific community, as confirmed by the wealth of studies published on the topic, and is important for having more successful therapeutic interventions in the future.

1.6   Flt3

The Flt3 protein is a RTK that belongs to the Pdgfr subfamily (also known as RTK III subclass family) together with Pdgfrα, Pdgfrβ, Kit and Fms (see table 1.1 on page 28). The existence of the gene translating for Flt3 was first discovered in mouse [187,188] and few years later verified in human as well [189,190]. The *flt3* gene is mapped to the 13th human chromosome (13q12) [191] and 24 exons are expressed with tissue-specific pattern, predominantly in hematopoietic progenitor cells in the bone marrow, thymus and lymph [189]. Secondary sites of expression are placenta, brain and gonads [192]. Furthermore, according to the cell expression profile, Flt3, is found mainly in early myeloid and lymphoid progenitors [193] with its tyrosine kinase activity involved in proliferation, differentiation and apoptosis of hematopoietic cells [194].

The 993 amino acids long protein can be found glycosylated (158-160 kiloDalton form) and non-glycosylated (130-143 kDa form) at the extracellular domain [195,196]. The glycosylation appears to maturate the protein with the consequent translocation on to the cell membrane.

The domain topology, shared by all the members of the III subclass of RTK, is composed by the extracellular ligand-binding moiety, with 5 immunoglobulin-like domains , a transmembrane portion predicted as α-helix and the cytosolic part with the juxtamembrane domain, the catalytic domain (with the kinase insert region) and the C-terminal tail.

The three-dimensional arrangement of Flt3 kinase domain is highly similar to the other PTKs as depicted by X-ray crystallography [117]. In addition, the available structure revealed an auto-inhibition mechanism mediated by the juxtamembrane domain, later confirmed for other two members of the same subfamily, Kit [73] and Fms [197].

Flt3 is physiologically activated by the binding of another protein to its extracellular portion. The so-called Flt3 ligand (FL) was first discovered in mice and later was

identified the human homologue [198, 199]. It is expressed in several isoforms [200]; the principal one is a 235 residues long transmembrane protein composed by an extracellular part (the ligand domain) and a small cytoplasmic one interconnected by a segment spanning the membrane [201]. The generation of the second form of the Flt3 ligand is achieved by protein cleavage with the release of the soluble extracellular domain [198,199].

With the binding of the receptor to its cognate ligand, Flt3 monomers are co-localized and their dimerization is triggered. RTK dimerization leads to conformational changes accompanied by concomitant release of the steric inhibition and activation of the kinase domains (described in subsection 1.4.1 on page 28). The stable activity of these catalytic moieties is achieved by transphosphorylation of cytosolic tyrosine residues. Thus the signal is transduced inside the cell and several biochemical pathways are initiated. After few minutes of activity the complex receptor-ligand is internalized and degraded [202].

It follows two examples of how Flt3 signal regulates cell growth and apoptosis.

Mapk pathway: the phospho-tyrosines at the intracellular part of Flt3 can recruit an adaptor protein, Grb2, that promotes the exchange of GDP with GTP in Ras protein [203,204]. This GTPase starts a kinase cascade involving Raf1, Erk1 and Erk2 [205]. Then, the mitogen-activated kinases (Erk1 and Erk2) translocate into the nucleus and activate transcription factors [206];

Pi3k pathway: the phosphoinositol 3-kinase activation appears to be mediated by adaptor proteins like Gab1, Gab2 or SHP2 which are phosphorylated by Flt3 [204]. Pi3k stimulates the biosynthesis of crucial regulatory proteins via kinase cascades including Pdk1, Pkb, mTor and S6k and inhibits cell apoptosis via phosphorylation of Bad, a protein which belong to the Bcl2 protein family [194].

Other signaling systems are likely to be downstream of the Flt3 physiological activation, such as Src family kinases and Jak/Stat kinases cascade. However, the role played by the receptor within those pathways is less clear [204].

The primary but not essential role of Flt3 in mammals has been revealed by the relatively normal haematopoiesis of Flt3 knockout mice [207] and by the large implication of the aberrant protein forms in leukemias [208]. About 10 years ago the first report of mutated Flt3 in hematopoietic malignancies was published [163]. Later efforts led to the identification of different oncogenic forms of the protein and their relevance in cancer (for examples [209–212]). Two major variations at the primary structure of Flt3 are frequently found in acute myeloid leukemia (AML), acute lymphoblastic leukemia (ALL) and myelodysplasia [194]. A sequence insertion at the juxtamembrane domain, called internal tandem duplication (ITD), characterizes the constitutively activated form of Flt3 in about 30% of AML patients [163] (further discussed in chapter 4 on page 79). The second malignant transformation of the protein can arise from different single point mutations within the KD [212].The more relevant one is the replacement of an aspartate at the beginning of the A-loop, predominantly by a tyrosine (Asp835Tyr) [213].

1.7   Methodology Survey

The development of computers with increased calculation power gave to the scientific investigations news possibilities in respect to data analysis as well as complex mathematical model building. In science, computers can be used to apply complex models to study different aspects of nature. They handle many numbers and elaborate mathematical functions, like the forecast and planets' trajectory predictions, and extend the human fantasy with the rendering at different size scales of the matter in different conditions.

Within this thesis, I largely used computational capabilities to study proteins and other molecules, their dynamics, interactions and to predict some of their behaviors. In this section some of the methods as well as their strengths and weaknesses, for addressing scientific questions, are described.

1.7.1   Sequence alignment

As discussed in section 1.1.2 on page 2, the analysis of the similarity between sequences of proteins of the same type from different species (paralogues) and proteins of different types but from the same specie (either orthologue or not) allowed a more complete and reliable mapping of the relationships among organisms in the life evolution. The concept behind these studies is straightforward: having the sequence of two proteins, the similarity comes from the pairwise comparison of their residues for position and type (for an example see figure 1.3.1 on page 9). Things get more complicated when the number of residues and of proteins, to be compared, increases. The issue is addressed by means of the combined usage of algorithms (reviewed in table 1.3) and computational power. A review about sequence alignment can be found in [214]. The algorithms have to "decide" whether two amino acids in different sequences can be matched or not. To handle the task, algorithms are provided with scoring tools for "judging" the similarity of the compared residues and the full alignment. The first scoring was based on residue *identity*. The alignment of amino acids scores 1 for identical aligned residues and 0 for diverse ones. Later, the *similarity* idea was applied by using scoring matrices based on the amino acid characteristics like physical-chemical properties, steric characteristics, and substitution frequencies found for homologue structures [219]. Thus, substitution tables are obtained by iterative comparison of subsequent residues in protein sequences. Optimal elongation of alignments of amino acid stretches can be obtained combining this informations with penalties for gap opening.

Table 1.3: *The most commonly used methods for aligning protein sequences.*

Algorithm	Principle	Method	Ref
Needleman-Wunsh	global pairwise alignment	dynamic programming	[215]
Smith-Waterman	local/global pairwise alignment	"	[216]
FASTA	multiple alignment	heuristic	[217]
BLAST	"	"	[218]

While applying these mathematical models to biology one should be aware that the optimal result does not mean error free. In fact, these models weight all matched amino acids of the investigated proteins in the same way, but under a biological point of view, not all residues have the same relevance. Depending on the type of the secondary, tertiary and quaternary structure of proteins, residues at different positions have different importance for structural requirements, like glycines for plasticity or proline for structure rigidity. Furthermore, in relation to the proteins' function, amino acids in certain positions are required for interaction with other molecules, like reagents of a reaction catalyzed by enzymes, ligands in case of receptors or other proteins for protein-protein interactions. Nature during evolution of species has selected amino acid mutations based on the residue role resulting in different degrees of conservation.

If structural informations are available for some of the proteins whose sequences are going to be analyzed, these should be used to judge and if possible to influence the final alignment. An examples is given in this chapter (figure 1.3.1 on page 9) where the structural alignment information of some proteins is used to guide the sequence alignment.

A step forward have been taken by using statistical modelling tools such as the hidden Markov models (HMMs) [220]. Computational programs are "trained" to recognize stretches of conserved sequences from multiple sequence alignments and develop statistical descriptors for specific amino acid patterns which are used to guide the new alignment.

1.7.2   Comparative modeling

When dealing with theoretical analysis of protein structures, the Cartesian coordinates of the atoms of these macromolecules are needed. Protein structures can be derived from experiments by collecting X-ray diffraction patterns of protein structures in crystals or by acquiring distance constrains between atoms from nuclear magnetic resonance (NMR). Nevertheless, experimental structures are not always available and thus alternative/complementary methods are needed to discover and predict proteins' behaviors.
Protein structures can be virtually created; one of the commonly used technique is the so-called comparative modeling [221]. This computational approach is based on the notion that the primary structure of proteins is conserved, through evolution, to a lesser extent than the higher level structures, namely secondary, tertiary and quaternary. In table 1.4, methods for comparative modeling are reported, for review [214, 221, 222].

Table 1.4: *The most commonly used methods for comparative modeling.*

Method	Ref	Program
Rigid bodies assembly	[223]	Swiss-model
Segment matching & coordinate reconstruction	[224]	Segmod
Spatial restraints satisfaction	[225]	Modeller

An amino acid sequence (target) can be modeled on the structure of a second protein (template) which are predicted to have the same folding. Based on the sequence alignment of the two proteins, the pairs of residues are spatially matched with the generation of the new coordinates for the target structure. Thus, the quality of the sequence alignment which determines the residues pairs is of primary importance. Usually, conserved regions, like secondary structure elements or patterns of residues implicated in the protein function, are identified in the structure of the template. Later, the alignment is optimized to match these conserved regions. The out-coming structure can be structurally refined with different protocols like energy minimization or simulated annealing. The resulting structure has to be checked for sterochemical quality, like ϕ and ψ angles distributions and bond lengths, angles etc., and for its feasibility of explaining already available biochemical data.

In addition, when the alignment reveals one or more long gaps, underlining structural variations between the two proteins, care must be taken on the structure generation. When new loops have to be built, meaning that the target sequence have non-correspondent stretches in the template, coordinates can be either assigned randomly and energy minimized or taken from experimentally known ones of other structures. The reliability of these additional loops depends on the length of these parts and the distance between the template extremities. The longer is the insertion, compared to the three-dimensional gap, the less reliable is the result.

In these studies, this approach was used in chapter 4 on page 79, to create protein structures for successive investigations.

1.7.3   Molecular dynamics

Molecular systems, where non-bonded interactions between atoms are present, possess intrinsic movements due to the changing distribution of their internal energy. Theoretical and empirical studies of proteins should take into account their dynamical behaviors. Movements of proteins are understood as a variety of different atomic dispositions which are specific for each protein system and are ruled by physical-chemical properties such as steric hindrance of side chains or attractive and repulsive charges. In general, this molecular conformational changes can be either little, with simple structure fluctuations due to the energy present at a given temperature within the system, or large as consequence of major modifications, such as phosphorylation of residue and binding of ligands.

Molecules can be described by mathematical models where the atomic positions, radii, masses and charges as well as the covalent bonds (length, angles) of their topologies are considered. Furthermore, the kinetic component is calculated from the velocities by mathematical integration of the equation of motion derived from the second Netwon's law:

$$F = m * a$$

Initial atomic velocities are used to start the compute of the kinetic component. Forces are then used to calculate the new atomic positions and velocities by integration of the equation of motion after a defined period of time (time step). The iteration of this cycle yield to the deterministic evolution (dependent from the previous steps) of the system

respect to the time. For reference [226].

The well known limitation of this method is how atoms are described. While using molecular mechanics (MM) model, the atoms of a simulated protein are described as balls with partial charges and the bonds are depicted as harmonic springs. The omission of all electrons speed up the calculation permitting longer time scale simulation [8] but decrease the accuracy of the system evolution.
Another issue of MD simulation is the lenght of the computed time life of a macromolecule. Certain biological phenomena concerning motions of proteins occur in a time scale which is not achievable by normal MD simulations.

The production of a trajectory usually involves three steps: the initialization of the system, its equilibration and production phase (see figure 1.13). During initialization ve-
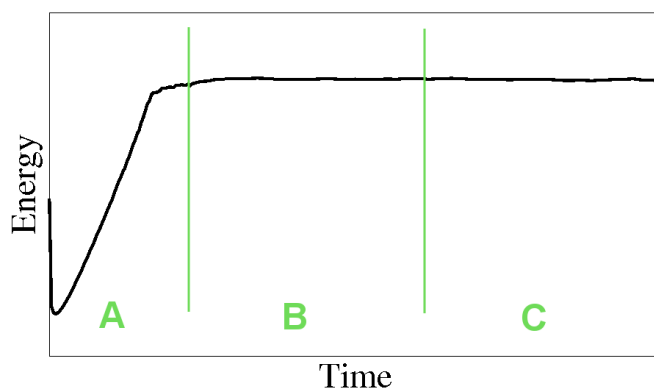

Figure 1.13: *Total energy profile in MD simulations. The progression of the total energy of the system is plotted against time. Different MD phases are identified and labeled in green. A) initialization phase, B) equilibration phase and C) production phase.*

locities are given to the atoms to calculate the first round of forces. When no velocities are available from a previous MD simulation, they are assigned randomly according to the Maxwell-Boltzmann distribution at given temperature. During equilibration the system is let evolve shortly to adjust velocities and to bring the system at the nearest thermic equilibrium an then the production phase.
Working with proteins some steps have to be added, this is due to the fact that these macromolecules are half way between liquid and solid state. In other words, the covalent bonds oscillations have to be restrained to reduce the number of degrees of freedom for the system. In the case that the solvent is wanted to be described explicitly in the trajectory, a certain number of water molecules have to added around the protein. The whole system needs to be energetically minimized to avoid bad steric contacts. Then a first round of MD is used to relax the solvent while the protein atoms are restrained in their initial positions. The next step consists in warming up the system, to the targeted temperature,

---

[8]To give an idea about the correlation between the real time and the simulated time of MD the general performance achieved with GROMACS-3.3.1 program package installed on a double processor Xeon 3.4 Ghertz 64 bit is reported. A system with approximately 40000 atoms needs 35 hours to calculate one nanosecond of trajectory in double precision, with particle mesh ewalds PME and a cut off of 0.9 nm for coulombic interactions.

i.e. 300 Kelvin, and to adjust the velocities. This is an important step for diminish the influence of the randomly assigned initial velocities in the final trajectory. The system is thus equilibrated for pressure and temperature using algorithms which every tot steps scale the velocities to match the set pressure and temperature within a given period of time. Eventually, the production phase is run and the system properties are collected for further analysis.

The reproducibility of this technique is an important issue because of the chaotic nature of multi-body dynamics. The several thousands particles affect the velocity of the single one by multiple interactions resulting in random trajectories. The word reproducibility is thus intended for averages of properties of the system calculated for relatively long simulations. Computational simulations of proteins should investigate a thermodynamic equilibrium of the system. The farther from the equilibrium the less reliable is the final trajectory. In chapters 4 on page 79 and 5 on page 107, MD is applied within some experiments.

1.8 Aims of the study

Being a study on the area of pharmaceutical sciences, the topic focuses on the etiology and the possible cure of specific diseases. In this thesis, the tyrosine kinase domain of TKs was investigated for the physiological behaviors and the pathological alterations found in certain cancers. The aims driving the investigations on this protein domain were essentially 2:
i) broaden the knowledge about the protein features such as sequences, structures and dynamics that are key elements for the molecular recognition between ligands and macromolecules and ii) build a framework where structure-based drug design can be applyied for putative ligands at the ATP-binding pocket.

Although the kinase domain has been largely studied under many aspects, some questions have not been answered yet. How do kinases achieve different conformations for their KD? Are there differences at the sequence level which can account for variations in their structures? Which similarities can be exploited in drug discovery and how can selectivity be achieved throughout the protein family? Answers were searched by the comparison of sequences, structures and dynamics of several kinase domains.

From a collaboration with clinicians working on one of the oncogenic tyrosine kinase, Flt3 bearing ITDs, questions about what is the molecular mechanism of Flt3 autoinhibition and of its disruption by ITDs were asked. To evaluate these issues the dynamics of the several constructs of the protein was simulated by MD in order to suggest hypotheses about oncogenic and normal activation.

Chapter 2

Tyrosine kinase drug discovery: what can be learned from solved crystal structures?

1

2.1 Abstract

Understanding molecular recognition during ligand binding is crucial for rational drug design. Protein kinases are a well known pharmacological target, especially in the therapeutic area of cancer. Solved crystal structures of tyrosine kinase domains in complex with ligands deliver insight into the binding event. These experimental data were collected and analyzed by means of molecular modelling techniques. Common molecular recognition patterns were depicted and studied among this class of proteins. The results of this analysis and the consequences for rational design of inhibitors are presented.

2.2 Introduction

Protein kinases are enzymes that transfer the γ-phosphate group of ATP to a peptidic substrate. These proteins are divided in two subfamilies based on the molecular target of the phosphorylation. Serine/threonine and tyrosine kinases transfer phosphates to hydroxyl groups of serine/threonine and tyrosine residues, respectively.
Tyrosine kinases are classified in receptor and non-receptor proteins, according to their function and localization in the cell, and sequence similarity [127]. Both are implicated in cell signalling: they do amplify, translate and integrate signals from inside and outside the cell [227]. Epidermal growth factor receptor (Egfr), platelet-derived growth factor receptor (Pdgfr), fibroblast growth factor receptor (Fgfr), vascular endothelial growth factor receptor (Vegfr) and insulin receptor (InsR) are examples showing the implications of these proteins in cell growth, differentiation and development. Deregulation of their catalytic activity is correlated to oncogenic and inflammatory processes as well as diabetes

---

[1]This chapter is already published as proceedings of an invited lecture; the reference can be found in [75]

and other diseases [39].

Several types of domains form a functional protein tyrosine kinase. Extracellular and transamembrane domains are present only in the case of receptors; while SH2, SH3 and other domains which up- and down-regulate the kinase catalytic activity (under physiological conditions) and interact with several cell components can occur in both classes. The conserved part of members of the protein kinase family is the catalytic domain, so-called kinase domain whose overall three-dimensional structure is depicted in figure 2.1 on the facing page.

The kinase domain is characterized by structural plasticity resulting in the existence of several conformations as shown by crystallographic studies [77]. The catalytic domain undergoes a conformational change to become enzymatically productive [68].

According to the catalytic requirements, the kinase domain has two recognition sites: one for the ATP (phosphoryl donor), placed in a cleft between the N- and C-terminal lobes and one for the peptide substrate (phosphoryl acceptor) which is situated within the C-terminal lobe in correspondence of the activation loop C-terminal end (figure 2.1 on the next page).

Tyrosine kinase inhibitors interfering with ATP binding are of interest for anticancer treatment. Glivec which is used as first line therapeutic for chronic myelogenous leukemia (CML) represents a successful example of this approach [228–230]. During the last years, a wealth of biochemical and structural data about the binding of these molecules has been produced. In fact, many ligand-kinase domain complexes have been crystallized and their three-dimensional structures have been solved by means of X-ray techniques. The coordinates of these complexes are publicly available at the Protein Data Bank (PDB, http://www.rcsb.org/pdb). In this manuscript we report the results of a comparison of these structures combined with multiple sequence alignments of tyrosine kinase domains. Common characteristics of the complexes are depicted and minimal requirements for binding deduced.

2.3 Results and Discussions

2.3.1 The Kinase domain

The overall three-dimensional structure of the kinase domain, depicted in figure 2.1 on the facing page, is conserved throughout the protein kinase family. The N-terminal lobe is a twisted β-sheet of five antiparallel β-strands and one α-helix, and the C-terminal lobe is composed by eight α-helices and four β-strands. These two lobes are connected by a segment called hinge region and they delimit the cleft representing the ATP-binding site. Other important moieties of this domain are: the nucleotide-binding loop, the αC helix, the catalytic and the activation loop (see figure 2.1 on the next page).

For easier understanding of the presented data, residue numbering and representations are referred to the sequence and crystal structures of the insulin receptor (PDB codes: 1IR3 [68] and 1IRK [67]).
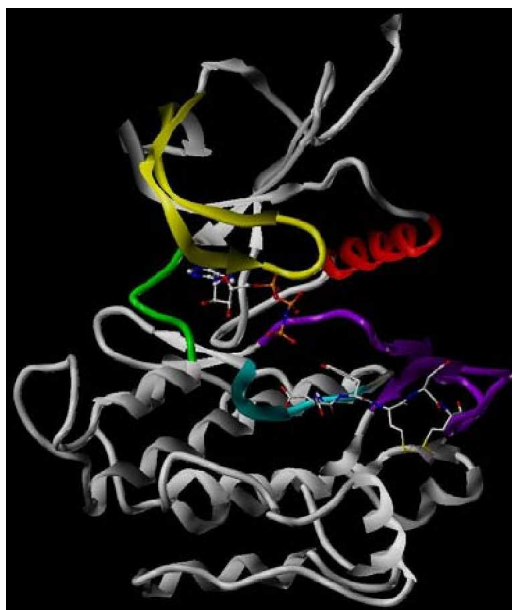
Figure 2.1: *The three-dimensional structure of the kinase domain. The tertiary structure of the kinase domain of the insulin receptor in complex with adenylyl imidodiphosphate (AMP-PNP), and peptide substrate (ball-and-stick and stick representations, respectively; PDB code: 1IR3) is depicted. Main regions are colored: nucleotide-binding loop (yellow), αC helix (red), hinge region (green), activation loop (violet) and catalytic loop (cyan).*

2.3.2 Variety of ligands

The PDB was screened and Cartesian coordinates of 44 structures of the catalytic domain of several tyrosine kinases in complex with ligands were retrieved (PDB entries are found in the supplementary information section). These ligands have IC_{50} values ranging from nM to μM [73, 82, 231–240].

The first analysis revealed the presence of 15 structures of kinase domains in complex with ATP-like molecules. Interesting, among these there are two kinase domains in complex with ATP analogue chemically linked with peptide that behaves as bi-substrate ligand.
The remaining 29 structures are kinase domains in complex with various types of inhibitors that belong to several chemical families (figure 2.2 on the following page).

The pyrimidine ring of the adenine which characterized ATP-like ligands is found as basic moiety in several groups of inhibitors: aminopyrimidine, pyrazolopyrimidine, pyridopyrimidine and quinazole derivatives.
Dihydroindolone and an isomer of it characterize other two classes of compounds. Staurosporine, which is a well-known ATP-competitive inhibitor with a wide range of action among protein kinases [237] and its analogues are dihydroisoindolone based molecules.
Recently, tyrosine kinase domains have been crystallized in complex with two aminoxazole derivatives. Finally, the last two inhibitors taken into consideration are a flavonol (quercetin) and an azepinone (debromohymenialdisine).
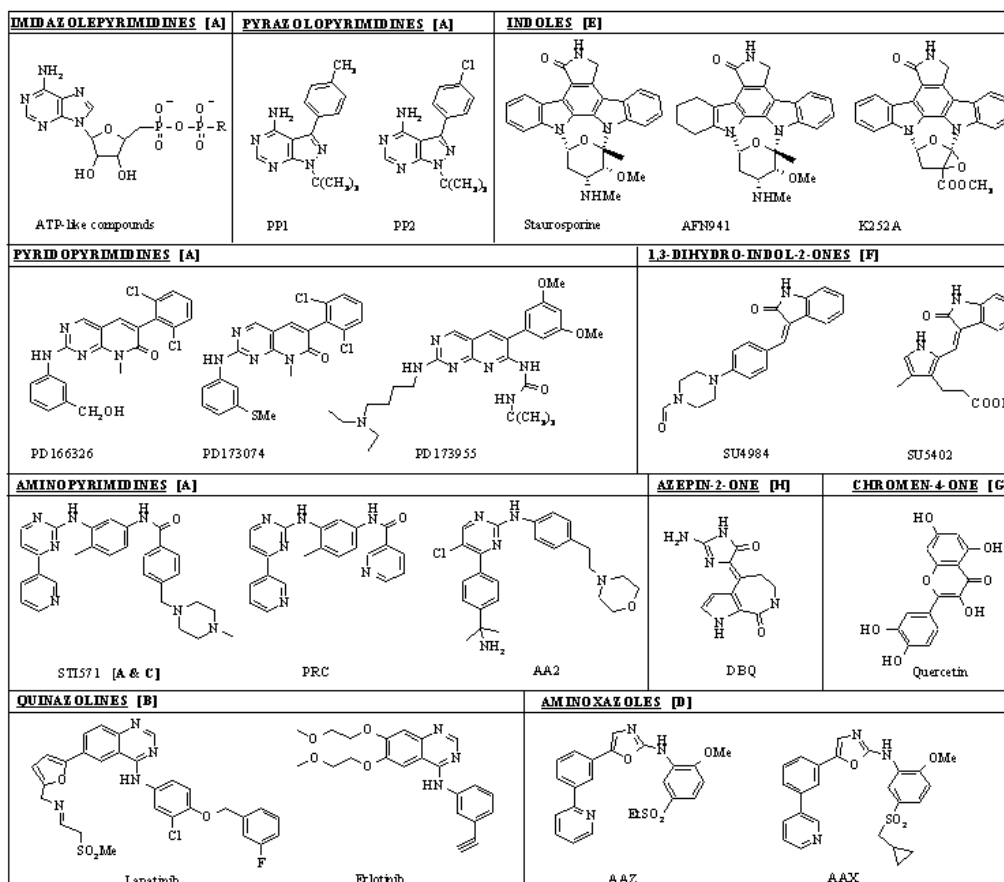
Figure 2.2: *Families of compounds. 2D-molecular structures of the ligands found in complex with the kinase domain. Names of compounds and families are reported. Annotations of the ligand cores are given in squared brackets (referring to figure 2.4 on page 48).*

In order to analyze the protein-ligand interactions computer-aided visualization of the complexes was used. The ligands interact with the kinase domain via hydrogen bonds to the hinge region and more or less extended hydrophobic interactions with the region expanding from the hinge towards the nucleotide-binding loop and/or the αC helix.

2.3.3 Common residues in the complexes

For each complex the residues interacting with the ligand were studied by examination of a protein area of 4.5 Å in the proximity of the ligand. The amino acids found were compared at the sequence level by means of multiple sequence alignment of all structures taken into account. This procedure allowed the highlighting of common residues involved in conserved interactions between the ligands and the kinase domains. In all complexes the common ligand-protein interaction pattern is constituted by hydrophobic interactions involving 4 residues and a single hydrogen bond.

These residues in the case of insulin receptor correspond to: Ala1028 (from the core of the N-terminal lobe), Met1139 (from the C-terminal lobe), Leu1002 and Val1010 of the nucleotide-binding loop and Met1079 of the hinge region (figure 2.3).
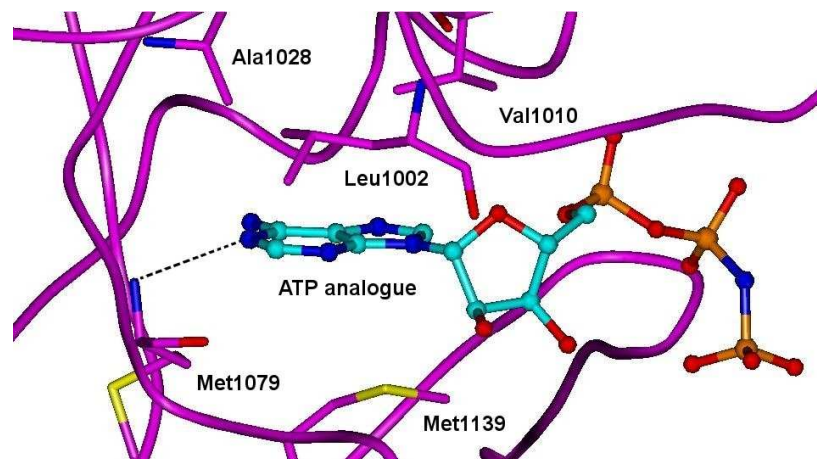
Figure 2.3: *Common interaction pattern in the ligand-kinase domain complexes. Secondary structure of the kinase domain of insulin receptor (1IR3) is represented as a tube (magenta), ATP analogue in ball-and-stick representation and the common residues as color-coded sticks (carbon in cyan for the ligand and magenta for the protein, oxygen in red, nitrogen in blue, sulphur in yellow and phosphorus in orange and the hydrogen bond as black dashed line).*

2.3.4 Defining and determining the ligand core

Superposition and visualization of the complexes permitted to depict in which part of the protein the conserved residues interacting with the ligand are localized. Common interactions are found at the same binding region in all structures, called adenine region in a previous work [74].

The moiety of the ligand interacting with this region was examined and defined as "ligand core". The results of this analysis are summarized in figure 2.4 on the following page.

It was generally possible to cluster ligand cores based on the ligand families with each family having a single ligand core. However, for STI-571, the ligand core was determined twice because of the two different modes of binding found in 4 complexes e.g. one involving the pyridine and the other the pyrimidine moiety (namely STI-571 bound to Abl (pdb codes: 1IEP [235] and 1OPJ [82]), Kit (1T46 [73]) and Syk (1XBB [239])).
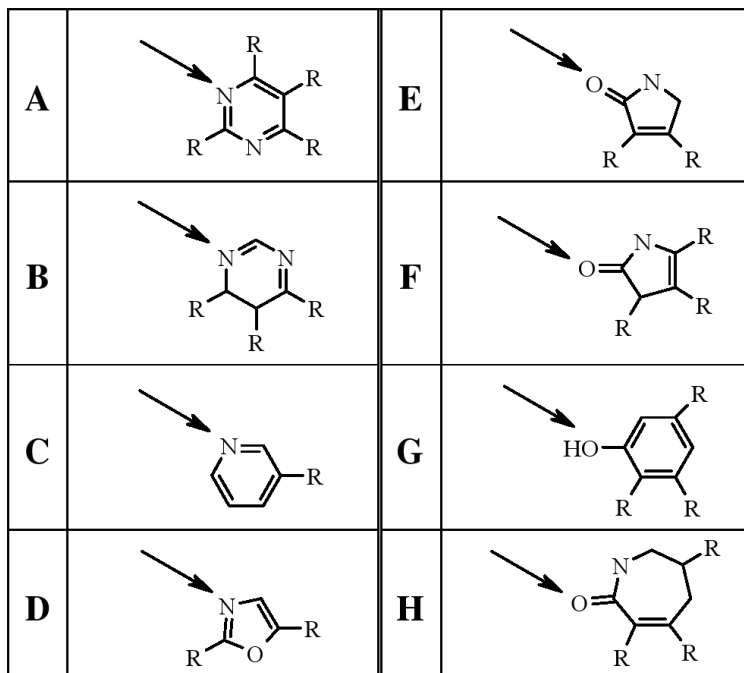
Figure 2.4: *Ligand cores. In the drawn structures, R is in accordance with the rest of the molecules as in figure 2.2 on page 46. Arrows indicate, for each ligand core, the atom involved in the hydrogen bond to the hinge region.*

2.3.5   Amino acid variability around the ligand core

For the 4 residues involved in hydrophobic interactions with the ligand cores, a consensus on the type of amino acid was expected. In fact, the hydrophobic character of these residues was checked by means of multiple sequence alignment (data not shown) of catalytic domains of 90 tyrosine kinases [38, 127]. The amino acids variability at the four positions was calculated based on the multiple sequence alignment and is reported in table 2.1 on the next page.

The table shows that at each position there is a rather dominant amino acid type. The alternative residues have always hydrophobic side chains, and thus the variability concerns only their bulkiness.

Table 2.1: *Amino acid variability of the hydrophobic pattern interacting with the ligand core.Variance (expressed in percentage) of the amino acid type at the positions (numbering referred to insulin receptor) involved in hydrophobic contacts with the ligand core. The data reported about the amino acid variability of the protein kinases [74] differ slightly in their values from the data here presented. This is most likely due to the used data sets, here only protein tyrosine kinases are considered which are a subfamily of the protein kinases. Nevertheless, the relative preference of amino acids shown here corresponds to the previous report [74].*

Leu1002	69%	Leu
	30%	Ile
	1%	Val
Val1010	96%	Val
	3%	Ile
	1%	Leu
Ala1028	92%	Ala
	4%	Ile & Val
	4%	Cys,Phe & Leu
Met1139	89%	Leu
	9%	Met
	2%	Val & Phe

2.3.6   Active/inactive: residues movements

The superposition of the apo and the ATP-bound crystal structures of the kinase domain of the insulin receptor revealed that a conformational change occurs upon ligand binding (see figure 2.5 on the following page). The N-terminal lobe is moved towards the second lobe when the ligand is bound, narrowing the cleft of the ATP-binding site.

To estimate the domain's closure around the ligand core, distances between amino acids in the N-terminal and C-terminal lobe were measured (reported in table 2.2). The selected residues are the ones involved in the common interactions: from the Cα of Met1139 (C-terminal lobe) to the Cα of Leu1002, Val1010 and Ala1028 (N-terminal lobe).

Distance variations (range 0.8-2.2 Å) shown in table 2.2 are meaningful of a conformational change between the productive and the unproductive structures. Whereas the variation of distances measured in other regions of these two complexes is significantly smaller (0.1-0.3 Å).
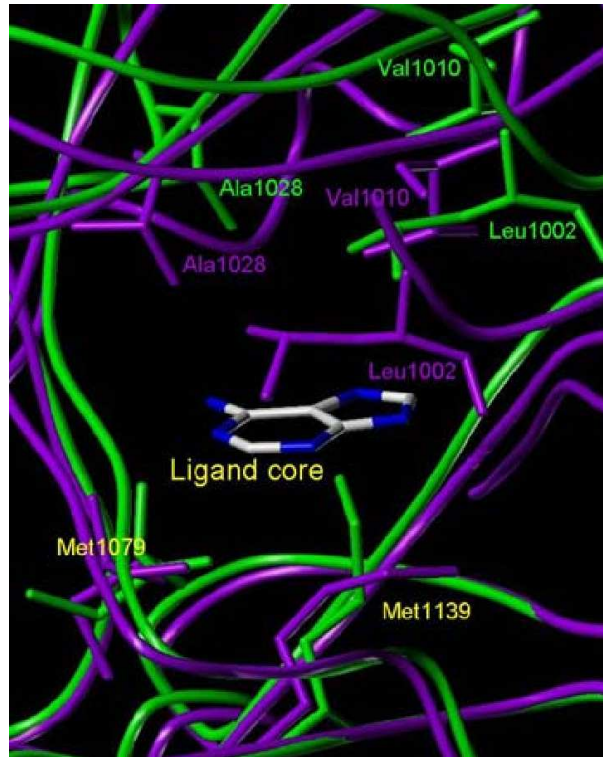
Figure 2.5: *Conformational changes around the ligand core. The superposition of active and non-active conformations of insulin receptor is depicted here. The protein main chain (tube representation) and the common residues (sticks representation) are colored in magenta (1IR3) and in green (1IRK). Adenine of the AMP-PNP is also drawn as sticks (in blue the nitrogen atoms and in light grey the carbon atoms), the pyrimidine moiety represents the ligand core.*

Table 2.2: *Distances between residues of the N-terminal lobe and Met1139. Distances in Angstrom from Cα of the Met1139 to Cα of the N-terminal lobe residues are reported. The inactive conformation is labeled "Inact" (resolution: 2.1 Å) and the active conformation "Act" (resolution: 1.9 Å).*

	Inact	**Act**	**Δ**
Ala1028	12.2	11.4	0.8
Leu1002	13.6	11.8	1.8
Val1010	15	12.8	2.2

2.4 Conclusion

In the present study, by the comparison of the structures of the tyrosine kinase domain in complex with ligands, we could show that binders of this protein family share a common feature that we called "ligand core". As a consequence, the interactions involving this part of the molecule were understood as minimal requirements for binding: a hydrogen bond to a main chain atom of the hinge region and apolar interactions with the side chain of 4 residues with conserved hydrophobic character.

This represents a reduction of the complexity for the pharmacophore compared to the adenine of the ATP that binds the hinge region via 2 hydrogen bonds.

The domain's closure upon ligand binding, calculated for the insulin receptor structures reveals that the ligand is likely to stabilize the protein conformation with the ligand core playing the major role. Other tyrosine kinase domains have been crystallized in their apo and ligated forms, and their conformational changes were checked by measurements of the distances earlier mentioned. The distance's variations are smaller and less appreciable than the ones reported for insulin receptor. This fact is due to several experimental aspects (size of ligand, presence of other domains interfering with the catalytic one, phosphorylation of the activation loop, conditions of crystallization, etc.) that might lead to differences in the degree of closure of the two lobes. This is in agreement with the general idea of protein "breath" [241].

In the year 2004 a similar study was published by Vulpetti and Bosotti [74]. The approach used and the considerations and point of view in this study and theirs are slightly different leading towards different and complementary results.

While we focused on defining the ligand core, the authors identified 38 residues of the catalytic domain of protein kinases which built up the ATP pocket. They studied the amino acid variance among the protein family focusing on how to achieve ligand binding selectivity. This work and the list of the ligand cores will extend with the number of the ligand-tyrosine kinase structure solved. The knowledge of the ligand cores might be useful in drug design when hypothetical ligand-binding modes are considered and suggested, for example while validating the results achieved from docking methods.

2.4.1 Supplementary information available

Table 2.3: *The table collects the PDB codes and other information of all tyrosine kinase complexes taken into consideration in the present study*

PDB code	kinase	ligand	year	PDB code	kinase	ligand	year
1IR3	InsR	ANP	1997	1OPJ	Abl	STI 571	2003
1GAG	InsR	ATP-pept	2001	1OPK	Abl	PD166326	2003
1I44	InsR	ACP	2001	1OPL	Abl	PD166326	2003
1RQQ	InsR	ATP-pept	2003	1XBB	SYK	STI 571	2004
1JQH	IGFR	ANP	2002	1XBC	SYK	Staurosporine	2004
1K3A	IGFR	ACP	2001	1U59	ZAP-70	Staurosporine	2004
1AGW	FGFR1	SU4984	1998	1BYG	CSK	Staurosporine	1999
1FGI	FGFR1	SU5402	1998	1QPJ	LCK	Staurosporine	2000
2FGI	FGFR1	PD173074	1999	1QPD	LCK	Staurosporine	2000
1OEC	FGFR1	AA2	2004	1QPC	LCK	ANP	2000
1M17	EGFR	Erlotinib	2002	1QPE	LCK	PP2	2000
1XKK	EGFR	Lapatinib	2004	1SNU	ITK	Staurosporine	2004
1Y6A	VEGFR2	AAZ	2005	1SM2	ITK	Staurosporine	2004
1Y6B	VEGFR2	AAX	2005	1AD5	HCK	ANP	1997
1PKG	KIT	ADP	2003	2HCK	HCK	quercetin	1997
1T46	KIT	STI 571	2004	1QCF	HCK	PP1	1999
1R0P	MET	K-252A	2003	1YVJ	JAK3	Stauro-anal	2005
1JPA	EPHR-B2	adenine	2001	1U54	ACK1	ACP	2004
1MQB	EPHR-A2	ANP	2003	1U4D	ACK1	DBQ	2004
1FPU	Abl	PRC	2000	2SRC	Src	ANP	1999
1IEP	Abl	STI 571	2001	1KSW	Src	ATP anal	2002
1M52	Abl	PD173955	2002	1MP8	FAK	ADP	2003

Chapter 3

Structural arrangement and ligand binding features by TK domains comparison

3.1 Abstract

In this chapter the collection of crystallographic structures of tyrosine kinase domains is reported as well as their multiple sequence alignment. Five newly defined clusters of residues of the KD are used to highlight conformational changes between different states of TKs. The way of arrangement of these clusters is used to classify the proteins in groups with similar dynamic behavior. Moreover, a pool of amino acids at the interface between the two lobes is identified as possible key feature responsible for structure disposition and dynamics of the KD.

Different ligand bindings are presented and their chemical characteristics are discussed as well as their pharmacological aspects. Finally, the issues of ligand conformational selection versus induced fit in the framework of tyrosine kinases is explained based on some examples.

3.2 Introduction

Protein tyrosine kinases (PTKs) constitute a large family of enzymes branched off the eukaryotic protein kinase superfamily during organisms evolution [51]. During the past years the crucial role of these macromolecules have been recognized in the normal physiology of animals [17]. On the other hand, their disregulation have been largely implicated with several diseases with a major component in tumors [39]. Based on the sequence identity, tyrosine kinases are classified in two groups, receptor tyrosine kinases (RTKs) and non-receptor tyrosine kinases (nRTKs), further by several subfamilies [52]. The multiple sequence alignment of PTKs reveals a common part, the kinase domain (KD), with 12 conserved subdomains [52], reviewed in chapter 1. During the last decades, this protein domain was intensively studied, under several aspects such as function, regulation, structure and dynamics. Important characteristics of the kinase domain were elucidated through structural investigations, as for example the spatial arrangements of non-productive and

53

productive KD states [67,68], the auto-inhibitory characters [115,117] and ligand-binding inhibition [235]. More than 100 crystallographic structures of the KD are publicly available and many more studies concerning other aspects can be found in the literature. Several definitions and classifications of KD features have been described, for examples the Gly-rich loop, A-loop, C-helix, DFG motif, etc., within the introductory chapter.

Despite the efforts, the issue concerning the role of single amino acid stretches in determining the stability of a privileged conformation is not completely solved. Here, an updated (March 2007) collection of crystallographic structures of kinase domains is reported and a new vision concerning structure and dynamics to further address the above mentioned issue, is given. With the comparison of these structures, 5 clusters of residues were defined as important for determining the state of the kinase. Their concerted movements and three-dimensional dispositions appeared to be shared between KD states from different PTKs. Furthermore, some characteristics of ligand-protein complex are revealed mainly based on the prototype of TK inhibitors: Imatinib.

3.3   Materials and Methods

3.3.1   Collection of information

138 structures of the catalytic domain of 30 unique tyrosine kinases were downloaded (March 2007) from the website of the protein databank (PDB) (at www.rcsb.org/pdb-/home/home.do). The PDB was screened in different ways for the sake of completeness:

- using the standard enzyme commission (EC) number [242]. Tyrosine kinases were first identified with E.C. 2.7.1.112 and later changed to 2.7.10.1 for RTKs and 2.7.10.2 for nRTKs. However, the annotation is not always respected as for example in the PDB files: 1JPA, 1QCF, 1IRK and 1PKG;

- via BLAST [218] search for sequence similarity. Using the primary sequence of a tyrosine kinase domain as query, the database can be screened and results are listed for similarity distance. Of course, the lower the similarity to the sequence reference the higher is the possibility that unrelated proteins are picked up;

- with the keyword "tyrosine kinase". But here, again the results depend on the annotations which are not always consistent. Furthermore, the structures of other parts of the TK proteins are found in abundance, like the SH2 domains.

The results obtained with these queries were then checked and compared, and only the consensus of the PDB entries was used for the study. The PDB entry codes and general descriptions of the KDs can be found in table 3.2 on page 74.

The reference publication for each structures were consulted, if available, for further specifications on the structures.

To collect the primary structures of tyrosine kinases, the EXPASY (Expert Protein Analysis System) proteomics server (at us.expasy.org/sprot/) was used. Entries were collected and downloaded from the website, using the single TK name as query.

3.3.2 Alignment of TKs' sequences

The amino acid sequences of the collected kinase domains were aligned by means of BODIL [61] and CLUSTALX [243] in a two-step procedure. Firstly, a 3D-based structural alignment of the crystallographic structures of the unique KD was performed using the VEERTA module of BODIL. Then, the obtained alignment was used as template in CLUSTALX to carry out the final multiple sequences alignment of the entire set of TK sequences (90). The full lengths sequences were added in several steps with the alignment and isolation of the catalytic domains after each step. The alignments were carried out with the Gonnet scoring matrix [244]. The penalty value 10 and 20 were applied for gap opening and gap extension, respectively. Final multiple sequence alignment was obtained after manual optimization of the loop regions.

3.3.3 Alignment of TKs' structures

For structural comparison of catalytic domains, two modules of BODIL software package were used namely the VEERTA and RMSD programs. The program VEERTA allows automated pairwise alignments of protein structures without initial information, like sequence alignments [61]. Thus, it was used for revealing general structural similarity between two KDs. To appreciate relative movements of protein portions, the RMSD module, carrying out the structural alignment of desired subsets of atoms, was used.

3.4 Results and Discussions

3.4.1 The KD features

Proteins experience thermodynamic movements due to the energy of the system at a given temperature: the "breath" of proteins [241]. The kinase domain displays also larger movements undergoing changes of the protein state, from "off" to "on" and vice versa. The comparison of the crystallographic structures reveals that the relative position of the two lobes, N- and C-lobe, varies depending on the state of the protein (productive vs. non-productive). The principal changes seem to be the result of rotation and bending around the interface of the two lobes.

These motions appear to be coupled with the A-loop element which can variate largely in its conformation. This element is supposed to interchange several conformations between two extreme positions, the "close" form (Abl-inactive like form), with a compact fold disturbing the center of reaction (as displayed in 1IRK for Insr), and the "open" form with an extended conformation (in 1IR3 for Insr). On the one hand, the open conformation, often stabilized by phosphorylation of A-loop tyrosine/s, is associate with the top of the enzymatic activity and is assumed to exist for all TKs. On the other hand, the close conformation is more variable and raises the question whether this Abl-inactive like form [70] is always achieved by all the members of TK protein family.

Smaller structural changes are revealed, at the N-lobe, by the relative positions of the C-helix with respect to the β-sheet composed of 5 anti-parallel strands. These motions are supposed to tune the activity via formation of interaction between two conserved residues: the lysine of the third strand and the glutamate from the C-helix. The glutamate serves

for the correct positioning of the lysine which catalytic role is to bind the cofactor, ATP.

Important movements are observed for the conserved DFG motif that goes from an Asp-out position in the non-productive state to an Asp-in position in the productive one. This is coupled with the A-loop motion being at the beginning of this element and serves to align its aspartate to chelate one of the counter-ion (Mg^{++}) of the cofactor. The phenylalanine position appeared to be fixed by the hydrophobic packing of residues at the C-terminus of the C-helix and residues from the C-lobe. These general mechanisms of the catalytic moiety seem to have evolved with different characteristics for the members of the protein family.

While the features described above have been reported already in the literature, the crystal structure comparison within this work aimed to depict further key elements, if any, that may be ruling the dynamics of TKs. This is an attempt to address the question whether there are stretches of amino acids favoring or disfavoring the acquiring of a given conformation e.g. the Abl-inactive like form.

The observation of the crystallographic structures to depict changes in the rearrangements of KD portions, guided the definition of five different hydrophobic clusters of amino acids[1] involved together in specific movements or functions.

Cluster 1: hydrophobic core 1 (H1), defined as the residues taking part of the hydrophobic core of the N-lobe; in cyan.

Cluster 2: hydrophobic core 2 (H2), residues of the hydrophobic core of the C-lobe; in blue.

Cluster 3: the phenylalanine of the DFG motif; in brown.

Cluster 4: the "binding-core" residues interacting via side chain with ligand-core (as defined in chapter 2 on page 43; in pink.)

Cluster 5: residues from the middle to the C-terminal of the inner face of the C-helix; in yellow.

Although, the residues belonging to the clusters are conserved in the hydrophobic character of the side chain throughout the analyzed kinases (see following alignment). Exceptions are also found; polar amino acids contribute to the hydrophobic clusters with their carbon skeleton and/or with neutralized ionic pairs (e.g. Glu1104, Lys1174 of Insr).

The question raised is whether there is a correlation between the hydrophobic clusters and the known dynamical behavior.

---

[1]In the following list the colors are related to the cluster identification in the figures of this chapter.

```
                      P-loop                                              C-helix                        N-Lobe <

         1010      1020      1030      1040      1050      1060      1070      1080      1090      1100

Insr   ..FPCSVYVPDEWEVSREKITLLRELGQGSFGMVYEGNARDIIKGEA..ETRVAVKTVNESASLRER..IEFLNEASVMKGFT.CHHVVRLLGVVSK..GQPTLVVME
Igf1r  ..SAADVYVPDEWEVAREKITMSRELGQGSFGMVYEGVAKGVVKDEP..ETRVAIKTVNEAASMRER..IEFLNEASVMKEFN.CHHVVRLLGVVSQ..GQPTLVIME
Musk   ..PLLLNPKLLSLEYPRNNIEYVRDIGEGAFGRVFQARAPGLLPYEP.FTM.VAVKMLKEEASADMQ..ADFQREAALMAEFD.NPNIVKLLGVCAV..GKPMCLLFE
EphA2  YEDPNQAVLKFTTEIHPSCVTRQKVIGAGEFGEVYKGMLKTSSGKKE...VPVAIKTLKAGYTEKQR..VDFLGEAGIMGQFS.HHNIIRLEGVISK..YKPMMIITE
EphA4  YEDPNQAVREFAKEIDASCIKIEKVIGVGEFGEVCSGRLKVP.GKRE...ICVAIKTLKAGYTDKQR..RDFLSEASIMGQFD.HPNIIHLEGVVTK..CKPVMIITE
EphB2  YEDPNEAVREFAKEIDISCVKIEQVIGAGEFGEVCSGHLKLP.GKRE...IFVAIKTLKSGYTEKQR..RDFLSEASIMGQFD.HPNVIHLEGVVTK..STPVMIITE
EphB3  YEDPNEAVREFAKEIDVSCVKIEEVIGAGEFGEVCRGRLKQP.GRRE...VFVAIKTLKVGYTERQR..RDFLSEASIMGQFD.HPNIIRLEGVVTK..SRPVMILTE
Erbb2  TPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVK...IPVAIKVLRENTSPKAN..KEILDEAYVMAGVG.SPYVSRLLGICLTS...TVQLVTQ
Fgfr1  ..SEYELPEDPRWELPRDRLVLGKPLGEGCFGQVVLAEAIGLDKDKPNRVTKVAVKMLKSDATEKDL..SDFLSEMEMMKMIGKHKNIINLLGACTQ..DGPLYVIVE
Fgfr2  ..SEYELPEDPKWEFPRDKLTLGKPLGEGCFGQVVMAEAVGIDKDKPKEAVTVAVKMLKDDATEKDL..SDLVSEMEMMKMIGKHKNIINLLGACTQ..DGPLYVIVE
Vgfr2  ..CERLPYDASKWEFPRDRLKLGKPLGRGAFGQVIEADAFGIDKTAT..CRTVAVKMLKEGATHSEH..RALMSELKILIHIGHHLNVVNLLGACTKP.GGPLMVIVE
Kit    ..DPTQLPYDHKWEFPRNRLSFGKTLGAGAFGKVVEATAYGLIKSDA..AMTVAVKMLKPSAHLTER..EALMSELKVLSYLGNHMNIVNLLGACTI..GGPTLVITE
Csf1r  ..DPTQLPYNEKWEFPRNNLQFGKTLGAGAFGKVVEATAFGLGKEDA..VLKVAVKMLKSTAHADEK..EALMSELKIMSHLGQHENIVNLLGACTH..GGPVLVITE
Flt3   ..DFREYEYDLKWEFPRENLEFGKVLGSGAFGKVMNATAYGISKTGV..SIQVAVKMLKEKADSSER..EALMSELKMMTQLGSHENIVNLLGACTL..SGPIYLIFE
Ret    ..DAFKILEDPKWEFPRKNLVLGKTLGEGEFGKVVKATAFHLKGRAG..YTTVAVKMLKENASPSEL..RDLLSEFNVLKQVN.HPHVIKLYGACSQ..DGPLLLIVE
Tie2   ..VKNNPDPTIYPVLDWNDIKFQDVIGEGNFGQVLKARIK....KDG.LRMDAAIKRMKEYASKDDH..RDFAGELEVLCKLGHHPNIINLLGACEH..RGYLYLAIE
Met    NPELVQAVQHVVIGPSSLIVHFNEVIGRGHFGCVYHGTLL.DNDGKK...IHCAVKSLNRITDIGEV..SQFLTEGIIMKDFS.HPNVLSLLGICLRS.EGSPLVVLP


Itk    ..PVTAGLRYGKWVIDPSELTFVQEIGSGQFGLVHLGYWLNKDK.......VAIKTIREGAMSEE....DFIEEAEVMMKLS.HPKLVQLYGVCLEQ.A.PICLVFE
Btk    .PSTAGLGYGSWEIDPKDLTFLKELGTGQFGVVKYGKWRGQYD.......VAIKMIKEGSMSED....EFIEEAKVMMNLS.HEKLVQLYGVCTKQ.R.PIFIITE
Csk    ..AAQDEFYRSGWALNMKELKLLQTIGKGEFGDVMLGDYRGN.K.....VAVKCIKN.DATAQ....AFLAEASVMTQLR.HSNLVQLLGVIVEE.KGGLYIVTE
Lck    ..KPQKPWWEDEWEVPRETLKLVERLGAGQFGEVWMGYYNGHTK.......VAVKSLKQGSMSPD....AFLAEANLMKQLQ.HQRLVRLYAVVTQE...PIYIITE
Hck    ..KPQKPWEKDAWEIPRESLKLEKKLGAGQFGEVWMATYNKHTK.......VAVKTMKPGSMSVE....AFLAEANVMKTLQ.HDKLVKLHAVVTKE..PIYIITE
Fyn    ..TDLSVKTKDVWEIPRESLQLIKRLGNGQFGEVWMGTWNGNTK.......VAIKTLKPGTMSPE....SFLEEAQIMKKLK.HDKLVQLYAVVSEE..PIYIVTE
Src    ..PQTQGLAKDAWEIPRESLRLEVKLGQGCFGEVWMGTWNGTTR.......VAIKTLKPGTMSPE....AFLQEAQVMKKLR.HEKLVQLYAVVSEE...PIYIVTE
Abl1   ..VYGVSPNYDKWEMERTDITMKHKLGGGQYGEVYEGVWKKYSLT.......VAVKTLKEDTMEVE....EFLKEAAVMKEIK.HPNLVQLLGVCTRE..PPFYIITE
Fak1   ..DTYTMPSTRDYEIQRERIELGRCIGEGQFGDVHQGIYMSPENPAL...AVAIKTCKNCTSDSVR..EKFLQEBALTMRQFD.HPHIVKLIGVITE...NPVWIIME
Ksyk   .ADPEEIRPKEVYLDRKLLTLEDKELGSGNFGTVKKGYYQMKKVVKT.....VAVKILKNEANDPALK.DELLAEANVMQQLD.NPYIVRMIGICEA...ESWMLVME
Zap70  .SDPEELKDKKLFLKRDNLLIADIELGCGNFGSVRQGVYRMRKKQID.....VAIKVLK.QGTEKADT.EEMMREAQIMHQLD.NPYIVRLIGVCQA...EALMLVME
Jak2   LGFSGAFEDRDPTQFEERHLKFLQQLGKGNFGSVEMCRYDPLQDNTG...EVVAVKKLQ...HSTEEHLRDFEREIEILKSLQ.HDNIVKYKGVCYSAGRRNLKLIME
Jak3   ..GAQLYACQDPTIFEERHLKYISQLGKGNFGSVELCRYDPLGDNTG...ALVAVKQLQ...HSGPDQQRDFQREIQILKALH.SDFIVKYRGVSYGPGRQSLRLVME
Ack1   GPAGEGPLQSLTCLIGEKDLRLLEKLGDGSFGVVRRGEWDAPSGKT....VSVAVKCLKPDVLSQPEAMDDFIREVNAMHSLD.HRNLIRLYGVVLTP...PMKMVTE

                       ━━━━                       ━━━                    ━━━             ━━━          ━━
                        I                          II                    III             IV           V
          ═══════>      ═══════>     ═══════>         ooooooooooooo     ══════>      ═══════>
            β1            β2           β3                 α1              β4             β5
```
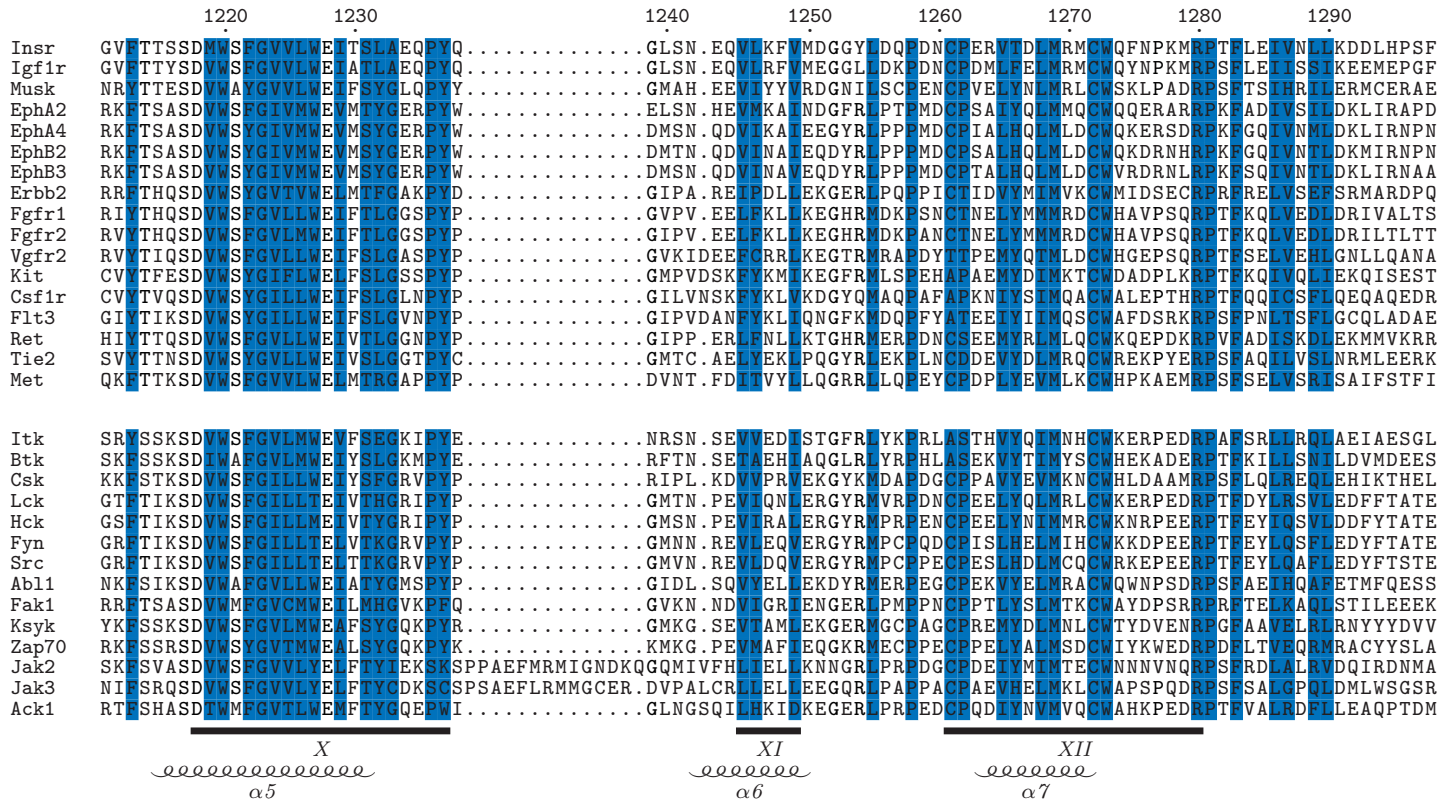
structural arrangement and ligand-binding features of TKs

hinge> C-Lobe   KID   *C-loop*   *A-loop*

```
        1110      1120      1130      1140      1150      1160      1170      1180      1190      1200      1210
Insr   LMAHGDLKSYLRSLRPEAENNPGRPPPTLQEMIQMAAEIADGMAYLNAKKFVHRDLAARNCMVAHDFTVKIGDFGMTRDIYETD...YYRKGGKGLLPVRWMAPESLKD
Igf1r  LMTRGDLKSYLRSLRPEMENNPVLAPPSLSKMIQMAGEIADGMAYLNANKFVHRDLAARNCMVAEDFTVKIGDFGMTRDIYETD...YYRKGGKGLLPVRWMSPESLKD
Musk   YMAYGDLNEFLRSMS..AQVSSPGPPLSCAEQLCIARQVAAGMAYLSERKFVHRDLATRNCLVGENMVVKIADFGLSRNIYSAD...YYKANENDAIPIRWMPPESIFY
EphA2  YMENGALDKFLREKD..........FSVLQLVGMLRGIAAGMKYLANMNYVHRDLAARNILVNSNLVCKVSDFGLSRVLEDDP.EATYTTS.GGKIPIRWTAPEAISY
EphA4  YMENGSLDAFLRKND..........FTVIQLVGMLRGIGSGMKYLSDMSYVHRDLAARNILVNSNLVCKVSDFGMSRVLEDDP.EAAYTTR.GGKIPIRWTAPEAIAY
EphB2  FMENGSLDSFLRQND..........FTVIQLVGMLRGIAAGMKYLADMNYVHRDLAARNILVNSNLVCKVSDFGLSRFLEDDTSDPTYTSALGGKIPIRWTAPEAIQY
EphB3  FMENCALDSFLRLND..........FTVIQLVGMLRGIAAGMKYLSEMNYVHRDLAARNILVNSNLVCKVSDFGLSRFLEDDPSDPTYTSSLGGKIPIRWTAPEAIAY
Erbb2  LMPYGCLLDHVRENR..........LGSQDLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLAR.LLDIDET.EYHAD.GGKVPIKWMALESILR
Fgfr1  YASKGNLREYLQARRYNPSHNPEEQLSSKDLVSCAYQVARGMEYLASKKCIHRDLAARNVLVTEDNVMKIADFGLARDIHHID...YYKKTTNGRLPVKWMAPEALFD
Fgfr2  YASKGDLREYLRARRYDINRVPEEQ.MTFKDLVSCTYQLARGMEYLASQKCIHRDLAARNVLVTENNVMKIADFGLARDINNID...YYKKTTNGRLPVKWMAPEALFD
Vgfr2  FCKFGNLSTYLRSKRKSLSDVEEEEFLTLEHLICYSFQVAKGMEFLASRKCIHRDLAARNILLSEKNVVKICDFGLARDIYKDP..DYVRKGDARLPLKWMAPETIFD
Kit    YCCYGDLLNFLRRKRGSYIERDVTPALDLEDLLSFSYQVAKGMAFLASKNCIHRDLAARNILLTHGRITKICDFGLARDIKNDS...NYVVKGNARLPVKWMAPESIFN
Csf1r  YCCYGDLLNFLRRKADSFSEQDLD.PLELRDLLHFSSQVAQGMAFLASKNCIHRDLAARNILLTNGHVAKIGDFGLARDIMNDS...NYIVKGNARLPVKWMAPESIFD
Flt3   YCCYGDLLNYLRSKRDEIEYENQKRVLTFEDLLCFAYQVAKGMEFLEFKSCVHRDLAARNILLTHGRITKICDFGLARDIMSDS...NYVVRGNARLPVKWMAPESLFE
Ret    YAKYGSLRGFLRESRSSLDHPDER.ALTMGDLISFAWQISQGMQYLAEMKLVHRDLAARNILVAEGRKMKISDFGLSRDVYEED...SYVKRSQGRIPVKWMAIESIFD
Tie2   YAPHGNLLDFLRKSR..IANSTAS.TLSSQQLLHFAADVARGMDYLSQKQFIHRDLAARNILVGENYVAKIADFGLSR...GQE..VYVKKTMGRLPVRWMAIESLNY
Met    YMKHGDLRNFIRNET..........PTVKDLIGFGLQVAKGMKYLASKKFVHRDLAARNCMLDEKFTVKVADFGLARDMYDKEYY.SVHNKTGAKLPVKWMALESLQT

Itk    FMEHGCLSDYLR.TQ..........FAAETLLGMCLDVCEGMAYLEEACVIHRDLAARNCLVGENQVIKVSDFGMTR.FVLDD...QYTSSTGTKFPVKWASPEVFSF
Btk    YMANGCLLNYLR.EM..........FQTQQLEMCKDVCEAMEYLESKQFLHRDLAARNVLVNDQGVVKVSDFGLTK.YVLDD...EYTSSVGSKFPVRWSPPEVLMY
Csk    YMAKGSLVDYLRSRG..........LGGDCLLKFSLDVCEAMEYLEGNNFVHRDLAARNVLVSEDNVAKVSDFGLTK.EASST...QDTG....KLPVKWTAEALRE
Lck    YMENGSLVDFLKTPS..........LTINKLLDMAAQIAEGMAFIEERNYIHRDLRAANILVSDTLSCKIADFGLAR.LIEDN...EYTAREGAKFPIKWTAPEAINY
Hck    FMAKGSLLDFLKSDE..........QPLPKLIDFSAQIAEGMAFIEQRNYIHRDLRAANILVSASLVCKIADFGLAR.VIEDN...EYTAREGAKFPIKWTAPEAINF
Fyn    YMNKGSLLDFLKDGE..........LKLPNLVDMAAQVAAGMAYIERMNYIHRDLRSANILVGNGLICKIADFGLAR.LIEDN...EYTARQGAKFPIKWTAPEAALY
Src    YMSKGSLLDFLKGET..........LRLPQLVDMAAQIASGMAYVERMNYVHRDLRAANILVGENLVCKVADFGLAR.LIEDN...EYTARQGAKFPIKWTAPEAALY
Abl1   FMTYGNLLDYLRECN..........VNAVVLLYMATQISSAMEYLEKKNFIHRDLAARNCLVGENHLVKVADFGLSR.LMTGD...TYTAHAGAKFPIKWTAPESLAY
Fak1   LCTLGELRSFLQVRK..........LDLASLILYAYQLSTALAYLESKRFVHRDIAARNVLVSSNDCVKLGDFGLSRYMEDST....YYKASKGKLPIKWMAPESINF
Ksyk   MAELGPLNKYLQQNR..........HVKDKNIIELVHQVSMGMKYLEESNFVHRDLAARNVLLVTQHYAKISDFGLSKALRADEN..YYKAQTHGKWPVKWYAPECINY
Zap70  MAGGGPLHKFLVGKR..........EIPVSNVAELLHQVSMGMKYLEEKNFVHRDLAARNVLLVNRHYAKISDFGLSKALGADDS...YYTARSAGKWPLKWYAPECINF
Jak2   YLPYGSLRDYLQKHK..........IDHIKLLQYTSQICKGMEYLGTKRYIHRDLATRNILVENENRVKIGDFGLTKVLPQDKE...YYKVKEPGESPIFWYAPESLTE
Jak3   YLPSGCLRDFLQRHR..........LDASRLILYSSQICKGMEYLGSRRCVHRDLAARNVLVESEAHVKIADFGLAKLLPLDKD...YYVVREPGQSPIFWYAPESLSD
Ack1   LAPLGSLLDRLRKHQ..........FLLGTLSRYAVQVAEGMGYLESKRFIHRDLAARNLLLATRDLVKIGDFGLMRALPQNDD..HYVMQEHRKVPFAWCAPESLKT
```

        V   VI   VII   VIII   IX
      ααααα   οοοοοοοοοοοοοοοοο   ⟹   ⟹   αααα
       α2   α3   β6   β7   α4

```
                1220        1230                      1240        1250        1260        1270        1280        1290
Insr    GVFTTSSDMWSFGVVLWEITSLAEQPYQ..............GLSN.EQVLKFVMDGGYLDQPDNCPERVTDLMRMCWQFNPKMRPTFLEIVNLLKDDLHPSF
Igf1r   GVFTTYSDVWSFGVVLWEIATLAEQPYQ..............GLSN.EQVLRFVMEGGLLDKPDNCPDMLFELMRMCWQYNPKMRPSFLEIISSIKEEMEPGF
Musk    NRYTTESDVWAYGVVLWEIFSYGLQPYY..............GMAH.EEVIYYVRDGNILSCPENCPVELYNLMRLCWSKLPADRPSFTSIHRILERMCERAE
EphA2   RKFTSASDVWSFGIVMWEVMTYGERPYW..............ELSN.HEVMKAINDGFRLPTPMDCPSAIYQLMMQCWQQERARRPKFADIVSILDKLIRAPD
EphA4   RKFTSASDVWSYGIVMWEVMSYGERPYW..............DMSN.QDVIKAIEEGYRLPPPMDCPIALHQLMLDCWQKERSDRPKFGQIVNMLDKLIRNPN
EphB2   RKFTSASDVWSYGIVMWEVMSYGERPYW..............DMTN.QDVINAIEQDYRLPPPMDCPSALHQLMLDCWQKDRNHRPKFGQIVNTLDKMIRNPN
EphB3   RKFTSASDVWSYGIVMWEVMSYGERPYW..............DMSN.QDVINAVEQDYRLPPPMDCPTALHQLMLDCWVRDRNLRPKFSQIVNTLDKLIRNAA
Erbb2   RRFTHQSDVWSYGVTVWELMTFGAKPYD..............GIPA.REIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDPQ
Fgfr1   RIYTHQSDVWSFGVLLWEIFTLGGSPYP..............GVPV.EELFKLLKEGHRMDKPSNCTNELYMMMRDCWHAVPSQRPTFKQLVEDLDRIVALTS
Fgfr2   RVYTHQSDVWSFGVLMWEIFTLGGSPYP..............GIPV.EELFKLLKEGHRMDKPANCTNELYMMMRDCWHAVPSQRPTFKQLVEDLDRILTLTT
Vgfr2   RVYTIQSDVWSFGVLMWEIFSLGASPYP..............GVKIDEEFCRRLKEGTRMRAPDYTTPEMYQTMLDCWHGEPSQRPTFSELVEHLGNLLQANA
Kit     CVYTFESDVWSYGIFLWEIFSLGSSPYP..............GMPVDSKFYKMIKEGFRMLSPEHAPAEMYDIMKTCWDADPLKRPTFKQIVQLIEKQISEST
Csf1r   CVYTVQSDVWSYGILLWEIFSLGLNPYP..............GILVNSKFYKLVKDGYQMAQPAFAPKNIYSIMQACWALEPTHRPTFQQICSFLQEQAQEDR
Flt3    GIYTIKSDVWSYGILLWEIFSLGVNPYP..............GIPVDANFYKLIQNGFKMDQPFYATEEIYIIMQSCWAFDSRKRPSFPNLTSFLGCQLADAE
Ret     HIYTTESDVWSYGVLLWEIVTLGGNPYP..............GIPP.ERLFNLLKTGHRMERPDNCSEEMYRLMLQCWKQEPDKRPVFADISKDLEKMMVKRR
Tie2    SVYTTNSDVWSYGVLLWEIVSLGGTPYC..............GMTC.AELYEKLPQGYRLEKPLNCDDEVYDLMRQCWREKPYERPSFAQILVSLNRMLEERK
Met     QKFTTKSDVWSFGVVLWELMTRGAPPYP..............DVNT.FDLTVYLLQGRRLLQPEYCPDPLYEVMLKCWHPKAEMRPSFSELVSRISAIFSTFI

Itk     SRYSSKSDVWSFGVLMWEVFSEGKIPYE..............NRSN.SEVVEDISTGFRLYKPRLASTHVYQIMNHCWKERPEDRPAFSRLLRQLAEIAESGL
Btk     SKFSSKSDIWAFGVLMWEIYSLGKMPYE..............RFTN.SETAEHIAQGLRLYRPHLASEKVYTIMYSCWHEKADERPTFKILLSNILDVMDEES
Csk     KKFSTKSDVWSFGILLTEIYSFGRVPYP..............RIPL.KDVVPRVEKGYKMDAPDGCPPAVYEVMKNCWHLDAAMRPSFLQLREQLEHIKTHEL
Lck     GTFTIKSDVWSFGILLTEIVTHGRIPYP..............GMTN.PEVIQNLERGYRMVRPDNCPEELYQLMRLCWKERPEDRPTFDYLRSVLEDFFTATE
Hck     GSFTIKSDVWSFGILLMEIVTYGRIPYP..............GMSN.PEVIRALERGYRMPRPENCPEELYNIMMRCWKNRPEERPTFEYIQSVLDDFYTATE
Fyn     GRFTIKSDVWSFGILLTELVTKGRVPYP..............GMNN.REVLEQVERGYRMPCPQDCPISLHELMIHCWKKDPEERPTFEYLQSFLEDYFTATE
Src     GRFTIKSDVWSFGILLTELTTKGRVPYP..............GMVN.REVLDQVERGYRMPCPPECPESLHDLMCQCWRKEPEERPTFEYLQAFLEDYFTSTE
Abl1    NKFSIKSDVWAFGVLLWEIATYGMSPYP..............GIDL.SQVYELLEKDYRMERPEGCPEKVYELMRACWQWNPSDRPSFAEIHQAFETMFQESS
Fak1    RRFTSASDVWMFGVCMWEILMHGVKPFQ..............GVKN.NDVIGRIENGERLPMPPNCPPTLYSLMTKCWAYDPSRRPRFTELKAQLSTILEEEK
Ksyk    YKFSSKSDVWSFGVLMWEAFSYGQKPYR..............GMKG.SEVTAMLEKGERMGCPAGCPREMYDLMNLCWTYDVENRPGFAAVELRLRNYYYDVV
Zap70   RKFSSRSDVWSYGVTMWEALSYGQKPYK..............KMKG.PEVMAFIEQGKRMECPPECPPELYALMSDCWIYKWEDRPDFLTVEQRMRACYYSLA
Jak2    SKFSVASDVWSFGVVLYELFTYIEKSKSPPAEFMRMIGNDKQGQMIVFHLIELLKNNGRLPRPDGCPDEIYMIMTECWNNNVNQRPSFRDLALRVDQIRDNMA
Jak3    NIFSRQSDVWSFGVVLYELFTYCDKSCSPSAEFLRMMGCER.DVPALCRLLELLEEGQRLPAPPACPAEVHELMKLCWAPSPQDRPSFSALGPQLDMLWSGSR
Ack1    RTFSHASDTWMFGVTLWEMFTYGQEPWI..............GLNGSQILHKIDKEGERLPRPEDCPQDIYNVMVQCWAHKPEDRPTFVALRDFLLEAQPTDM

                        X                           XI                  XII
                       α5                           α6                  α7
```

Figure 3.1: *Multiple sequence alignment of the KDs investigated. An extract of the alignment obtained as described in materials and methods is reported with single-letter code for amino acids. Just the sequence of kinases whose structure was experimentally resolved is shown. Names of the kinases are shown on the left and numbering is according to the Insr sequence (list and pdb codes in table 3.2 on page 74). Regions of the alignment are color-coded to identify the clusters defined previously and the catalytic residues. Cyan for cluster 1, blue for cluster 2, brown for cluster 3, pink for cluster 4, yellow for cluster 5 and red for catalytic residues. While mainly hydrophobic residues are found in the cluster regions, polar amino acids also contribute to it with their carbon skeleton. On the top, the known regions are reported and at the bottom the secondary structure assignment of Insr structure, 1IRK, is shown. Furthermore, the 12 subdomain as defined in [52] are found below the sequences (roman numbers).*

The "Insr-like" behavior

The dynamics of the kinases that belong to this group, are inferred from the crystallographic structures of Insr KD of the non-productive and productive state, 1IRK and 1IR3, respectively. In the basal conditions the non-productive conformation of the enzyme (see figure 3.2a on the next page) is maintained by the close form of the A-loop, which is stabilized by a network of hydrogen bonds (Hbonds) [92] and interferes with the binding of substrate and cofactor. The DFG motif is in the Asp-out conformation with its phenylalanine sandwiched between the binding-core residues. This interaction disturbs the binding of the ligand-core and is likely to restrain the uncoupled movements of the two lobes.

When activated the Insr undergoes specific changes of the 5 clusters (see figure 3.2b on the facing page). The phosphorylation of the tyrosines at the A-loop stabilizes its open form. The DFG motif switches into the Asp-in conformation with the phenylalanine trapped under the C-terminal of the C-helix by hydrophobic interactions. The C-helix swings down and toward the catalytic center bringing its glutamate to interact with and positioning the lysine of the third strand.

This movement is permitted by the rearrangement of the van der Waals contacts between the inner face of the helix and the β-sheet. Furthermore, the interaction of the adenine ring (the ligand-core) of ATP to the binding-core as well as the hydrophobic interactions of the phenylalanine of the DFG motif fixes the relative position of the two lobes favoring the catalysis.

Although, in absence of ligands a higher stability of the non-productive form, compared to the productive one, was obtained [245], the crystal structures revealed that the binding of ligands results in lower B-values for atoms in the productive conformation than in the non-productive [245]. Thus, the minor fluctuations of the two lobes might be necessary to trigger the phosphoryl-transfer by favoring the correct positioning of substrate, cofactor and catalytic side chains, with the consequent decrease of the entropic term of the reaction.

The comparison of X-ray structures of Kit (in figure 3.2 on the next page) depicts similar features for the transition between the inactive to the active form. Thus, the Insr dynamical behavior is expected also for the members of the Pdgfr family. However, the fine tuning of the movements may vary because residues in C-terminal position to DFG-motif and the Gly-rich loop interact in a different manner for Kit KD compared to Insr. A further difference of the two kinases is the small change of the C-helix position with the interaction between the catalytic glutamate and lysine which is already formed in the non-productive Kit (see figure 3.2c on the facing page). This suggests that the productive conformation could be closer to the non-productive one compared to Insr. Considering the presence of the juxtamembrane domain, whose role is to inhibit the enzyme [246] (see on chapter 4 on page 79 for details), it is conceivable that the C-helix movement of Kit can be different than Insr. While, in Insr, part of the regulation of the catalytic efficiency may be achieved via extended conformational changes, the presence of an "external" regulatory factor such as the juxtamembrane domain allows a reduction of the conformational changes needed for maintaining the control of the catalytic efficiency.

(a) *non-productive Insr (1IRK)*

(b) *productive Insr (1IR3)*

(c) *non-productive Kit (1T45)*

(d) *productive Kit (1PKG)*
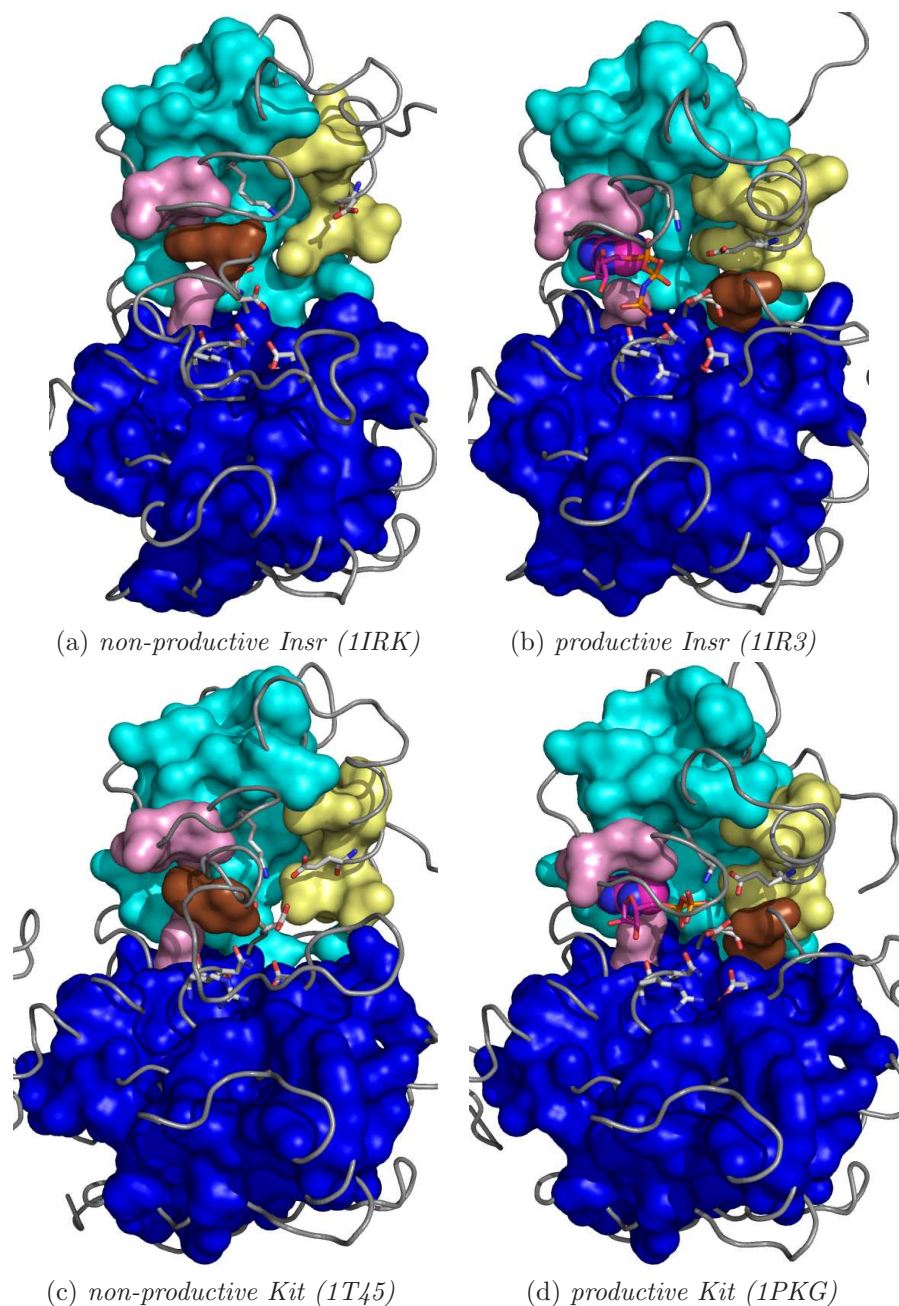
Figure 3.2: *Non-productive and productive states of the tyrosine kinases having "Insr-like" behavior. The crystallographic structures are identified by the PDB code in between round brackets. The protein chains are shown as gray ribbon and specific residue clusters are represented with the external molecular surface and color coded. In cyan the hydrophobic regions 1 (H1), in blue the hydrophobic regions 2 (H2), in pink the hydrophobic residues interacting with the ligand cores (as defined in [75]), in yellow the main hydrophobic residues of the C-helix and in brown the phenylalanine residue of the DFG-motif. The atoms and bonds are represented as sticks for the residues involved in the enzyme catalysis and for the ligands. The ligand-core atoms are shown in CPK. All represented atoms are color-coded: nitrogen in blue, oxygen in red, chlorine in green, phosphorus in orange and carbon in gray for protein atoms and magenta for ligand-atoms.*

The "Cdk-like" behavior

The first deviation from the normal mode ("Insr-like") is so far reported for two tyrosine kinases, Egfr and Csk. The term "Cdk-like" reveals the need of a second protein portion, interacting with the C-helix, for enzyme catalysis, as described for Cdks [111]. This idea was recently suggested by a study on Egfr showing crystallographic dimer formation in agreement with an increase of activity due to high local concentration of KD [72]. Despite the presence of the ATP mimicking molecule, the conformation in the PDB file 2GS7, depicted in figure 3.3a on the facing page, is likely to be representative for the non-productive conformation where the A-loop folds in an intermediate conformation and the position of the phenylalanine of the DFG motif is not yet fixed below it, as in the "Insr-like" mode. In contrast the structure in which the formation of a crystallographic dimer has been observed (2GS2) is likely to represent the productive state. The second monomer interacts with the outer face of the C-helix pushing this element closer to the reaction center with the consequent formation of the glutamate-lysine pair (see figure 3.3b on the next page).

Although, it may be argued whether Egfr is a TK representative case for the Cdk-like behavior, because the crystallographic results may be partially due to methodological artifacts, the stabilization of the active conformation by an external help for achieving the catalytic conformation was already known for Csk [172]. The non-productive Csk in complex with Staurosporine has a distort conformation of the C-helix with the glutamate separated from the lysine by the phenylalanine of the DFG motif (see figure 3.3c on the facing page). However, this might not be the only non-productive state but rather an intermediate configuration trapped by the ligand. In the active conformation of the KD of Csk (figure 3.3d on the next page) the linker between SH2 and SH3 domains interacts with the external surface of the C-helix which assume a position in which the essential catalytic glutamate-lysine pair is formed.

The "Src-like" behavior

The characteristics of Src kinase domain are likely to be extendible to the proteins of the same family, namely Src, Lck, Fgr, Fyn, Yes1, Blk, Hck and Lyn, as judged by the comparison of the available crystallographic structures of the family members. In addition to the catalytic moiety these proteins are composed by SH2 and SH3 domains, which structural arrangement have been suggested for KD auto-inhibition [247]. Thus, the apo/non-productive conformation deviates from the normal arrangements of the 5 defined protein portions, as depicted in figure 3.4a on page 64, with the major change in the closer configuration of the two lobes. The more compact structure of the inactive state leaves no space between the lobes for the C-helix to reach the catalytic center of the protein. The phenylalanine of the DFG motif is packed, as in the productive conformation, between the C-helix and the C-lobe upper face (with the respect of the view in the figures) and the dynamics of the A-loop appears to be restricted to an intermediary conformation. Until the recent publication of Src in complex with Imatinib [70], it was thought that Src-family elements would not be able to achieve a closer conformation of A-loop and the Asp-out form of the DFG-motif under physiological conditions. This conclusion was drawn because the displacement of the inhibitory domains from the auto-inhibited form, increases the catalytic activity [247, 248] leading to the phosphorylation of the tyrosine

(a) *non-productive Egfr (2GS7)* (b) *productive Egfr (2GS2)*

(c) *non-productive Csk (1BYG)* (d) *productive Csk (1K9A)*
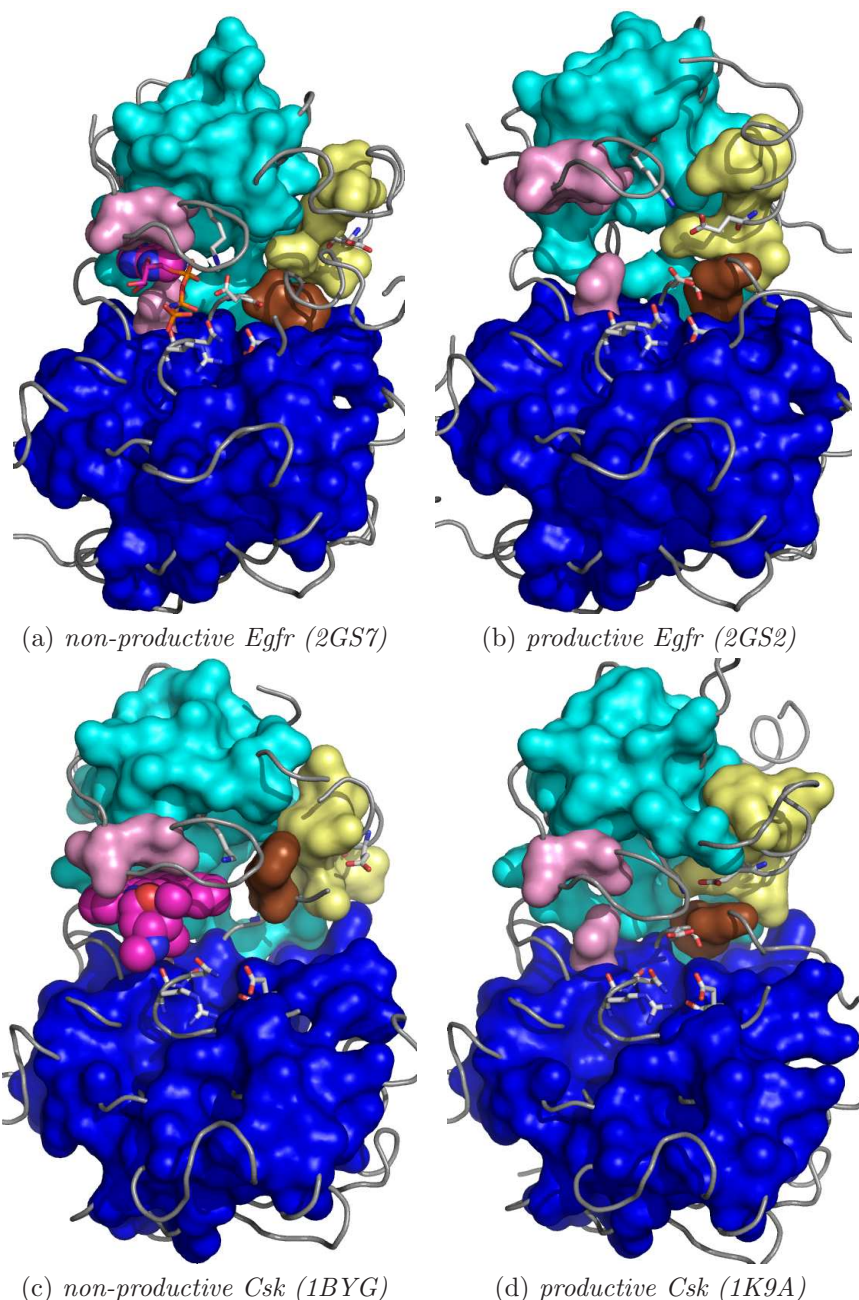
Figure 3.3: *Non-productive and productive states of the tyrosine kinases having "Cdk-like" behavior. The crystallographic structures are identified by the PDB code in between round brackets. The protein chains are shown as gray ribbon and specific residue clusters are represented with the external molecular surface and color coded. In cyan the hydrophobic regions 1 (H1), in blue the hydrophobic regions 2 (H2), in pink the hydrophobic residues interacting with the ligand cores (as defined in [75]), in yellow the main hydrophobic residues of the C-helix and in brown the phenylalanine residue of the DFG-motif. The atoms and bonds are represented as sticks for the residues involved in the enzyme catalysis and for the ligands. The latter has the ligand-core atoms in CPK representation. All represented atoms are color-coded: nitrogen in blue, oxygen in red, chlorine in green, phosphorus in orange and carbon in gray for protein atoms and magenta for ligand-atoms.*
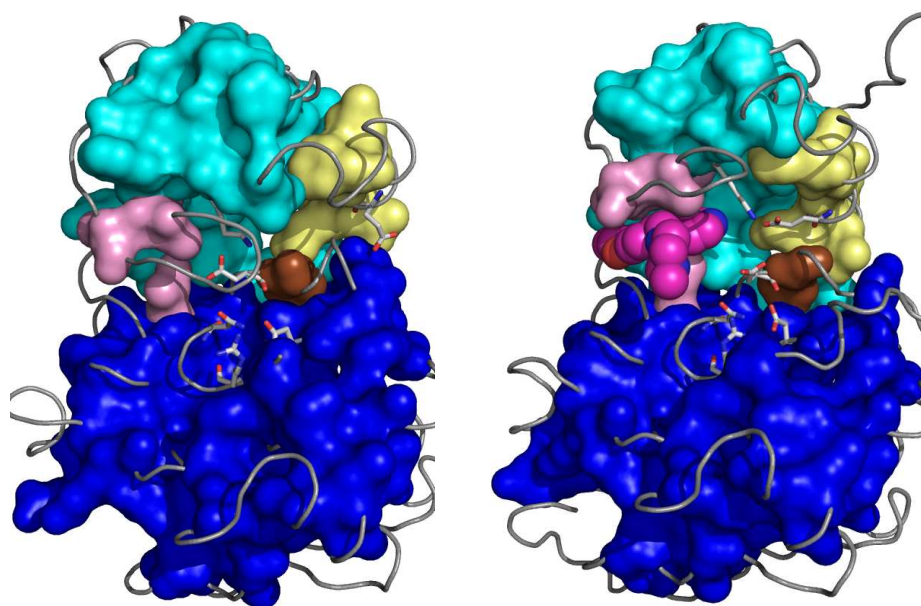
(a) non-productive form of Src (1FMK)

(b) bound-active form of Src (1Y57)
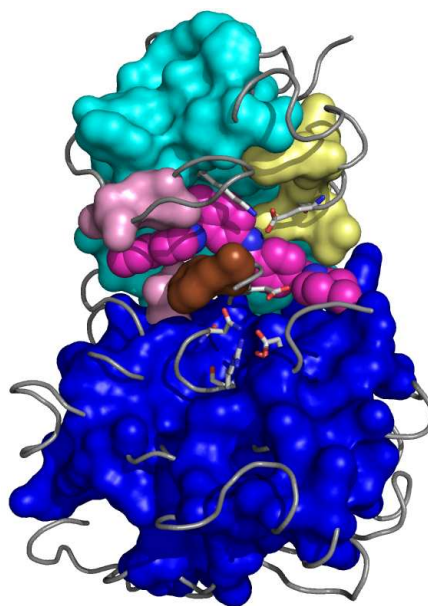
(c) bound-inactive form of Src (2OIQ)

Figure 3.4: *Apo-form and bound-forms of Src. The crystallographic structures are identified by the PDB code in between round brackets. The protein chains are shown as gray ribbon and specific residue clusters are represented with the external molecular surface and color coded. In cyan the hydrophobic regions 1, in blue the hydrophobic regions 2, in pink the hydrophobic residues interacting with the ligand cores (as defined in [75]), in yellow the main hydrophobic residues of the C-helix and in brown the phenylalanine residue of the DFG-motif. The atoms and bonds are represented as sticks for the residues involved in the enzyme catalysis. The ligand is depicted in CPK. All represented atoms are color-coded: nitrogen in blue, oxygen in red, chlorine in green, phosphorus in orange and carbon in gray for protein atoms and magenta for ligand-atoms.*

residue of the A-loop and the stabilization of the active enzyme.

Conformations similar to 1FMK are seen for Hck structures in complex with ligands (PDB codes: 1AD5, 2HCK, etc.) where a larger distance between H1 and H2 is achieved for the presence of the interaction between the ligand-core and the binding-core.

The comparison of the non-productive (figure 3.4a on the preceding page) with the productive state (figure 3.4b on the facing page) of the KD confirms the proper conformation taken by the release of the SH2 and SH3 domains for protein catalysis. Beside the opening of the two lobes, the major structural variation, is the swing of the C-helix positioning its glutamate for binding the lysine. A similar productive state conformation has been seen for Fyn (PDB code: 2DQ7), Lck (3LCK, etc.) and Hck (2HK5, etc.).

Interestingly, an arrangement of the 5 clusters comparable to the one of non-productive form of the kinases with "Insr-like" behavior has been observed by Src-family elements while binding ligands favoring this less populated conformation [70]. As demonstrated by the crystallographic structure of Src bound to Imatinib (PDB code: 2OIQ) (in figure 3.4c on the preceding page) and Lck with furano derivatives (2OFV and 2OG8), the ligand spans the length between the surfaces of the two lobes interacting at the binding-core and in the space usually filled by the phenylalanine of the DFG motif in the active conformation. In the resulting state, the displaced phenylalanine extends the interaction of the C-lobe part of the interaction-core with the ligand-core.

A possible TK behavior

Based on the observation of the apo-structures of Vegfr and Syk and on pharmacological results of Ret, a question is raised whether some tyrosine kinases may exhibit an "always-active" behavior. These TKs have all the crucial elements always in place for fostering the reaction despite the absence of substrates, co-substrate and phosphotyrosines. Vegfr and Ret KDs are classified in this class because no major changes are present in different crystallographic results independently from the phosphorylation state. Their experimental structures have very similar conformation with small changes in the case of Vegfr and tiny ones for Ret. The apo form of Vegfr (PDB code: 1VR2, depicted in figure 3.5b on the following page) shows that all residues involved in the catalysis are placed correctly for supporting the reaction. All the 5 clusters of residues are in productive position, only the phenylalanine of the DFG can change conformation as seen in the other crystal structures of the same kinase in complex with ligands. The extreme example is reported in figure 3.5a on the next page. The DFG motif is in the Asp-out conformation allowing the ligand to bind between the two lobes in an Imatinib-like manner. The phenylalanine packs against some of the residues of the binding-core cluster and the whole conformation appears a bit more tighter than the apo form. A second example (here not reported) is found in the PDB file 1YWN where the A-loop is phosphorylated and it is in a sort of close conformation, which seems to be induced by the ligand.

The catalytic efficacy is 10-fold enhanced by the phosphorylation of the tyrosine residue at the A-loop [151] which is supposed to further stabilize the conformation of this element (not visible in the crystallographic structure of the apo form).

Recently, four Ret structures (not shown) were published in complex with different ligands and in different state of phosphorylation [158]. The comparison of phosphorylated and non-phosphorylated states reveal the same arrangements of the whole protein which, as in Vegfr, appears ready for catalysis. The apo form is not there for telling us about

(a) *ligand-inhibited form Vegfr (2P2I)*

(b) *apo-form Vegfr (1VR2)*
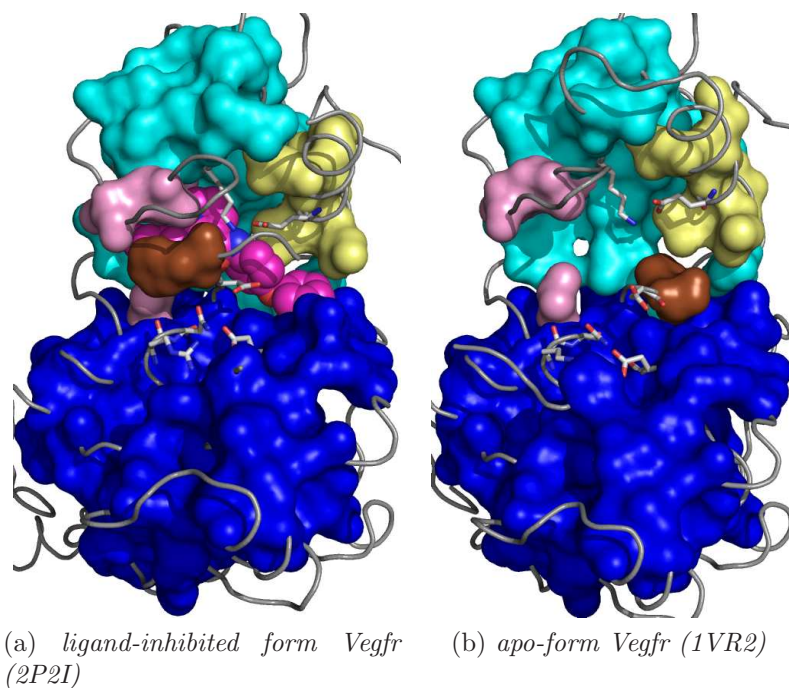
Figure 3.5: *Apo-form and bound-form of Vegfr. The crystallographic structures are identified by the PDB code in between round brackets. The protein chains are shown as gray ribbon and specific residue clusters are represented with the external molecular surface and color coded. In cyan the hydrophobic regions 1 (H1), in blue the hydrophobic regions 2 (H2), in pink the hydrophobic residues interacting with the ligand cores (as defined in [75]), in yellow the main hydrophobic residues of the C-helix and in brown the phenylalanine residue of the DFG-motif. The atoms and bonds are represented as sticks for the residues involved in the enzyme catalysis and for the ligands. The ligand is shown in CPK. All represented atoms are color-coded: nitrogen in blue, oxygen in red, chlorine in green, phosphorus in orange and carbon in gray for protein atoms and magenta for ligand-atoms.*

structural changes of this form. However, small changes (an increment of 3-4 folds) of catalytic efficacy are reported for the phosphorylated state [158]. The fact that the catalytic efficiency is increased in the phosphorylated state (stable productive form) may also reflect the observation that, in the non-phosphorylated state the productive form is in equilibrium with a less populated but existent non-productive one. Thus, further studies are needed before the category of the always-active can be clearly defined.

Structural considerations: the H1-H2 interface

As stated before, the movements of the N-lobe and C-lobe leading to a change of state in some kinase domains are understood as rotations, bending and sliding, which appeared independent for one hydrophobic core with respect to the other one. This partial freedom in the dynamics of the two lobes, prompted the analysis of the interface between H1 and H2 aimed to identify the putative mechanism for KD movements. Thus, based on the Insr crystallographic structures, 6 residues from H1 and 8 from H2 were identified and investigated (reported in table 3.1 on page 70).

The visual inspection of the two faced regions (figure 3.6 on the next page and figure 3.7 on page 69), formed either by residues from H1 or H2, revealed predominant van der Waals contacts and two polar interactions.

While the first electrostatic link is the hydrogen bond between the main chain atoms of Val1087 (H1) and Ile1175 (H2) the second one is the side chain interaction between a glutamate, Glu1104 (H1), and a lysine, Lys1174 (H2). The latter two residues are highly conserved throughout the tyrosine kinase family and this bonding is present in a large set of the experimental structures. Residues from H1 and H2 form two planes, observing the interface from the protein inside (figure 3.6 on the next page), and sort of pawls looking the back face (figure 3.7 on page 69). In details, Lys1174 and Tyr1149 from H2, both well conserved, embed H1 by holding Cys1083, Val1086 and Val1087. At the position of the valine residues, in Insr, the other TKs sequences reveal a maintained hydrobophic character by frequent substitutions to either leucine or isoleucine. Cys1083, often substituted by histidine in other TKs, seats on the Cα-atom of Gly1146. Taking a closer look, glycine at this position might be needed to support the movements of H1 without sterical hindrance. This is confirmed by the high degree of conservation of the glycine. Within the alignment, the amino acid is exchanged with alanine in few cases. Furthermore, the Cα of alanine is in plane with the interface, as observed on Abl structures.

Two variations of the interaction pattern for the defined interface are found comparing the non-productive and the productive structures of Insr (1IRK and 1IR3, respectively). The first, regards a single bond rotation of the side chain of Met1104 (from the hinge region) that results in a minor loss of van der Waals contact with Val1167 and Lys1174 (carbon chains). The second, a larger one, involves the change of packing among the residues, Phe1081, Cys1083, Val1086, Leu1150, Phe1155 and Ile1175, with major rearrangements for residues at position: 1081, 1150 and 1155 (see figure 3.6 on the next page). A switch of the conformation for the latter residues is frequently seen, although not always, between productive and non-productive states depicted by the experimental structures of TKs.

In a plausible scenario, the movements of the N- and C-lobe have their focus at this interface with changes of the hydrophobic contacts and the conservation of the two polar
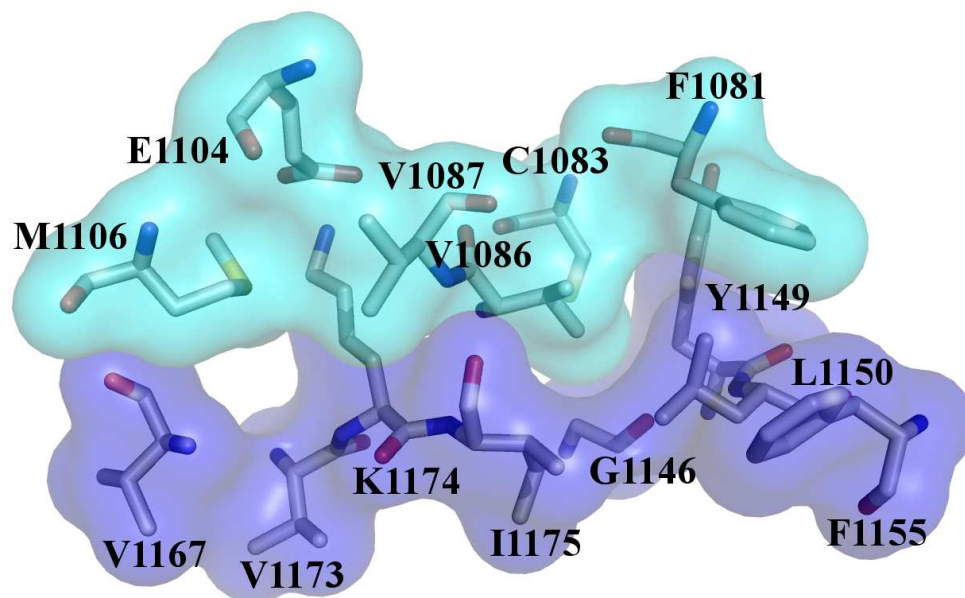
(a) *non-productive*
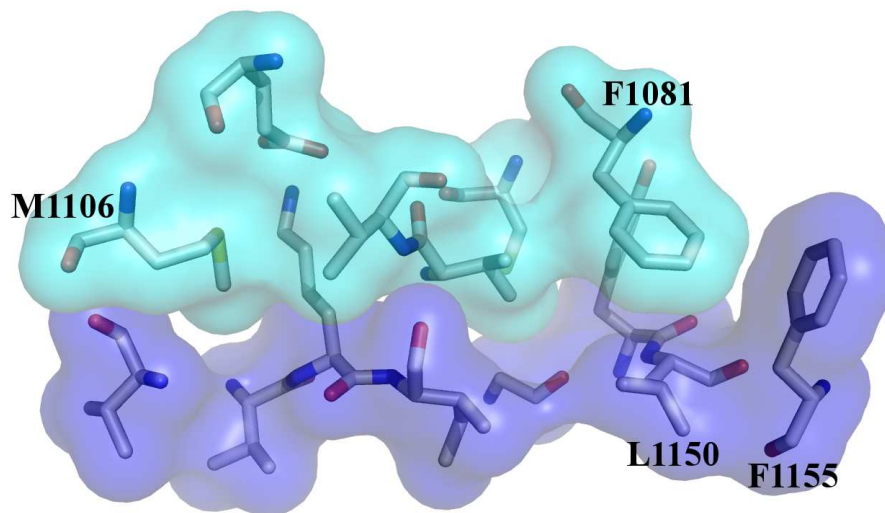
(b) *productive*

Figure 3.6: *H1-H2 interfaces for Insr non-productive (a) and productive states (b), view 1. The residues found at the interfaces in the two crystallographic structures are represented in sticks, labelled and color-coded for atoms: gray for carbon, red for oxygen, blue for nitrogen and yellow for sulphur. The surfaces of the atoms are also represented and color-coded for clusters, cyan for H1 and blue for H2. The amino acids labeled in (b) experience the highest side chain movements while the others are superimposable.*
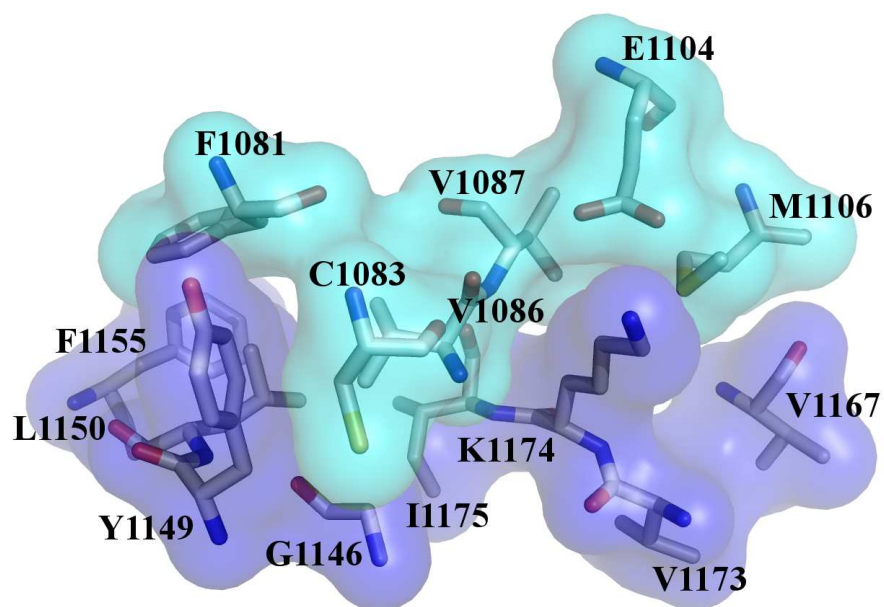
(a) *non-productive*
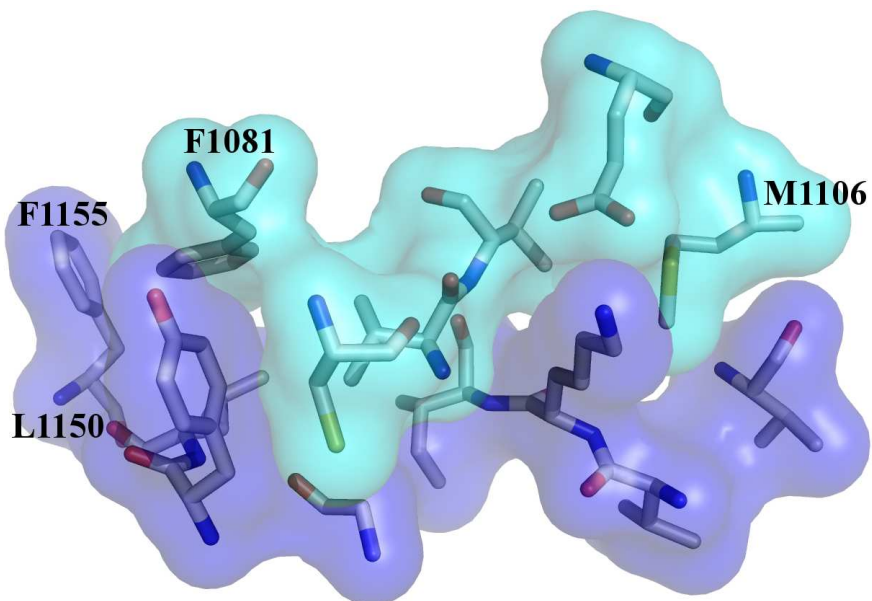
(b) *productive*

Figure 3.7: *H1-H2 interfaces for Insr non-productive (a) and productive states (b), view 2. The view is rotate 180 degree compare to figure 3.6 on the facing page. The residues found at the interfaces in the two crystallographic structures are represented in sticks, labelled and color-coded for atoms, gray for carbon, red for oxygen, blue for nitrogen and yellow for sulphur. The surfaces of the atoms are also represented and color-coded for clusters, cyan for H1 and blue for H2.*

Table 3.1: *Residues found at the interface of the two lobes and belonging to H1 and their interaction partners in H2 are reported. Numbering according to the fasta sequence and, in between brackets, the correspondent from the 1IRK PDB file.*

N-lobe residues	C-lobe partners
Phe1081 (1054)	Phe1155 (1128)
	Leu1150 (1123)
	Tyr1149 (1122)
Cys1083 (1056)	Gly1146 (1114)
	Tyr1149 (1122)
Val1086 (1050)	Gly1146 (1114)
	Ile1175 (1148)
Val1087 (1060)	Ile1175 (1148)
	Lys1174 (1147)
Glu1104 (1077)	Lys1174 (1147)
Met1106 (1079)	Val1167 (1140)
	Lys1174 (1147)

interactions. Thus, the ideal centroid would have intermediate coordinates between the two electrostatic bonds. However, major rearrangements of this area cannot be expected. Being the fulcrum, small variations, in the region, account for larger ones at the far extremities of the protein structure.

The hypothesis is indirectly proved by the highly conserved nature of the residues of the H1 and H2 interface. Interestingly, the variability of such positions is appreciated within the family of TKs and with a lower extent among elements of same subfamily, which might account for the diversity at the level of the dynamical behavior for their kinase domains.

The importance of the amino acids at this interface could be addressed by mutagenesis studies. One of the most interesting aspects to test would be the polar interaction between the conserved lysine and glutamate by their replacements with residues whose side chains are either shorter or bearing different charges. Furthermore, the stability and other consequent features of the whole protein might deviate, from the one observed for Insr, when the bulkiness of the lateral chain of the hydrophobic residues, like Phe1081, Cys1083, Val1086, Val1087, Met1106, Leu1150, Phe1155 and Ile1175 , is changed.

Ligand-binding features for KD

After the achievement of the minimal requirements for a molecule to bind the KD, identified in the ligand-core (in chapter 2 on page 43), further features of the ligand-KD complexes were searched by comparison of crystallographic structures. The case of Abl tyrosine kinase in complex with several ligands is here reported to explain a logical evolution of a putative TK ligand towards better pharmacodynamic properties.

From the ligand-core moiety additional interactions can be acquired by enlarging van der Waals contacts with the inner surface of the H1 as depicted in figure 3.8a and figure 3.8b on the next page. The place is often filled by an aromatic ring carrying diverse substitutions of one or more hydrogens with bulkier atoms to increase the hydrophobic complementarity with the protein (see figure 3.8a and figure 3.8b on the following page). The resulting molecule would draw into the protein driven by the hydrophobic effect, pushing the separation of the two lobes and getting trapped by the steric hindrance of the decorated ring.

This type of ligands stabilize the protein in a conformation where the glutamate-lysine pair is formed and the DFG can be either in the Asp-out form (in figure 3.8a on the next page) or Asp-in, with the phenylalanine fixed below the C-helix when the A-loop is phosphorylated (figure 3.8b on the following page).

Further interactions might be achieved with the elongation of the molecule. The longer ligand can reach the space between the C-helix and the C-lobe with the displacement of the phenylalanine of the DFG-motif, as depicted in figure 3.8c on the next page. Although, the major type of engagements are van der Waals contacts, in the complex ligand-protein polar interactions are also present. Beside the bond with the hinge region, described for the ligand-core in chapter 2 on page 43, Imatinib, as an example of ligand, binds two side chains, the hydroxyl group of the threonine known as the "gate-keeper" residue and the carboxyl group of the conserved glutamate from the C-helix. Furthermore, other three polar contacts involve main chain carbonyl groups, two of which of residues of the catalytic loop.

While developing TK inhibitors, the plasticity of these enzymes pushes the investigators to deal with different protein conformations and their lifetime, in agreement with the energetic landscape. Questions like: which is the most populated conformation? is this the best target? is the ligand binding due to conformational selection? or does also induced fit occur?, need to be answered to ameliorate the treatment of diseases, such as cancers. Some of these issues can already be addressed with published data.

Conformational selection of ligands was suggested by the comparison of *in-vitro* activities of two inhibitors towards Abl kinase domain. PD173955, crystallized with Abl in 1M52 (in figure 3.8a on the next page), results a more potent inhibitor than Imatinib with an IC_{50} of 5 nM and 100 nM, respectively [235]. The better potency of the former was appointed because it inhibits, with equal potency, both phosphorylated and non-phosphorylated form, while a worse activity of Imatinib is displayed against the phosphorylated form of Abl [235].

Furthermore, Imatinib has been shown to bind with a similar activity Kit, selecting the same protein conformation that is also depicted for the apo form [73]. It binds Syk TK with 2 order of magnitude lower K_i values, both phosphorylated and non-phosphorylated forms [239]. With the latter kinase, the confirmation of inhibitory activity linked to the conformation selection is given by the crystallographic study with the inhibitor (PDB code 1XBB) [239]. Syk presents the same conformation for the apo form and bound to the ligand, the A-loop is open and the DFG-motif is in the Asp-in form. In the complex, Imatinib is in a folded conformation, it binds with a different moiety as ligand-core, than in the case of Abl, and does not reach the site between the C-helix and the C-lobe. Thus, the X-ray investigation of this complex revealed a second configuration of the inhibitor, different than the extended one depicted within Abl and Kit.
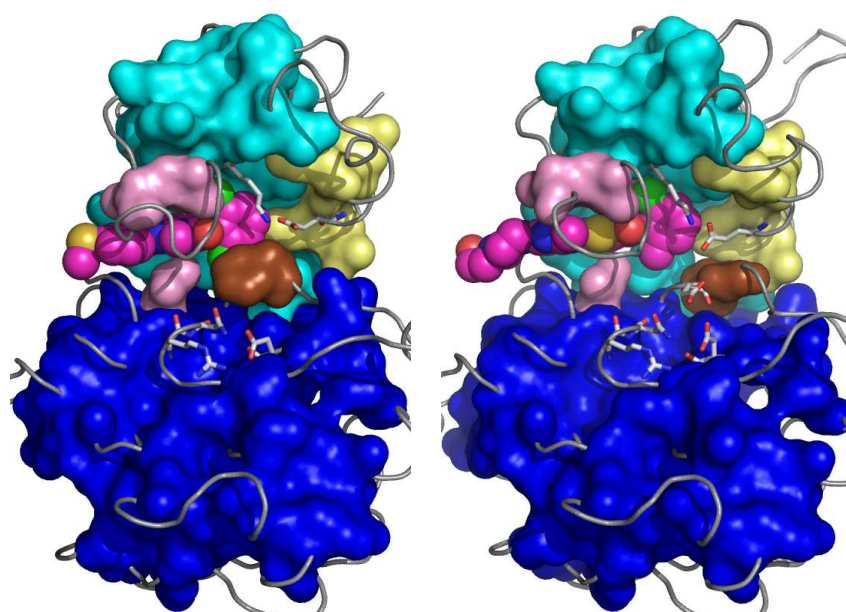
(a) *bound-active Asp-out form of Abl (1M52)*

(b) *bound-active Asp-in form of Abl (2GQG)*
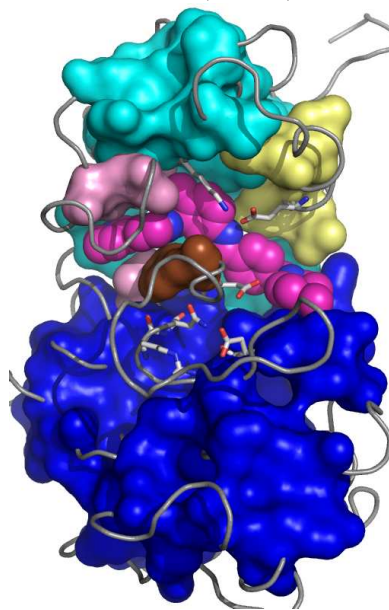
(c) *bound-inactive Asp-out form of Abl (1IEP)*

Figure 3.8: *Unbound and bound-forms of Abl. The crystallographic structures are identified by the PDB code in between round brackets. The protein chains are shown as gray ribbon and specific residue clusters are represented with the external molecular surface and color coded. In cyan the hydrophobic regions 1 (H1), in blue the hydrophobic regions 2 (H2), in pink the hydrophobic residues interacting with the ligand cores (as defined in [75]), in yellow the main hydrophobic residues of the C-helix and in brown the phenylalanine residue of the DFG-motif. The atoms and bonds are represented as sticks for the residues involved in the enzyme catalysis. The ligand atoms are in CPK. All represented atoms are color-coded: nitrogen in blue, oxygen in red, chlorine in green, phosphorus in orange and carbon in gray for protein atoms and magenta for ligand-atoms.*

As already anticipated, a fourth complex Imatinib-TK has been recently published under the PDB entry 2OIQ. The compound is depicted into Src KD, in figure 3.4c on page 64, with the same binding mode as in complex with Abl and Kit. The affinity of the compound against Src is one order of magnitude lower than against Syk and the binding is achieved 50 times slower than with Abl [70]. This is a clear example where the ligand induces changes in the conformation of the protein, but this conformational changes cost a thermodynamic penalty for the complex [70].

The last issue is dedicated to the persistence of ligand into the binding site which is related to the affinity of compound for the protein. The experimental value to quantify the interaction can be divided in the enthalpic and entropic terms; while the latter accounts for the amount of conformational populations of both binders, the first reflects their complementarity. Both terms can be optimized on the ligand side, rigidification and decoration of compounds are often used in medicinal chemistry. Here, as proof of the principle, we report the case of Imatinib optimization with the addition of a chlorine atom (Puttini et al., 2007; submitted). Testing the resulting compound, a higher affinity than the parent molecule could be demonstrated with *in-vitro* and *in-vivo* assays. Furthermore, wash-out experiments reveal that the better pharmacological profile can be attributed to the long-lasting effect of the designed compound.

This is a nice example on how the inhibition constant (K_i) of a compound can be ameliorate by lowering its off-rate kinetic parameter (κ_{off}).

3.5   Conclusions

90 proteins of the tyrosine kinase family are present in humans and many of them have already revealed their disfunctions linked to severe diseases. Indeed, Imatinib showed that some of these disfunctions can be cured effectively by xenobiotics targeting the kinase domain of TKs [181]. This success geared up the scientific community to investigate the kinase domain as possible target for several diseases. But nature has evolved this domain with a myriad of different mechanisms within the protein family resulting in a variety of dynamical behaviors. Thus, the question about these enzymes is whether there are amino acids stretches, in the sequence, which can affect the overall protein structure favoring certain conformations. If so, the resulting more populated conformations would be ideal targets for inhibitors because of their long existence. One third of the KD structures of tyrosine kinases is known by experiments, but sometimes only one conformation is present.

In tyrosine kinase field, we have examples of ligand conformational selection as well as induced fit, but the latter one appear to cost a thermodynamic penalty resulting in a lower inhibition efficacy.

The aim of this study was the isolation of stretches of residues in the kinase domain to directly infer the dynamical behavior of a certain tyrosine kinase in their sequences. In other words, the idea was to consider the kinase domain as unique fold with its deviations expressed in the different structures and dynamics of TKs. Thus, knowing the difference in the sequence, the one in the structure would be the consequence. This would help to guide the rational drug discovery approach for selection of the target to investigate.

Although a wealth of information about the kinase domain is published, the isolation of such key residues is not straightforward. The step taken, with the present work, was

the identification of groups of residues, close in space, which define regions of the protein that move together during its dynamical evolution. In fact, 5 clusters have been defined, namely the hydrophobic core of N-lobe (H1), the one of the C-lobe (H2), the interaction-core, the phenylalanine of the DFG-motif, and part of the C-helix. Their rearrangements between different states of the KDs seems to characterize four families, namely "Insr-like", "Cdk-like", "Src-like" and a possible "always-active".

The interface between two of the clusters revealed 14 quite conserved residues which interact with two polar and many apolar interactions. The particular configuration of this region is suggested to play an important role in KD structure and dynamics. Further investigations on how H1-H2 interface influences the stability and dynamics of the catalytic domain are reported on chapter 4 on page 79.

With the availability of more experimentally derived structures of KDs, further comparisons of the residues at the interfaces of these clusters might be indicative for similar behaviors.

3.6   appendix

Table 3.2: *Collection of the PDB files (March 2007)*

PDB code	kinase	ATP-ligand	domains	A-loop
1IRK	Insr	-	KD	closed ; no
1IR3	"	ANP	KD	open ; yes x3
1GAG	"	ATP-pept	KD	open ; yes x3
1I44	"	ACP	KD	intermed,dis. ; no
1RQQ	"	ATP-pept	KD+SH2-APS	open ; yes x2
1P14	"	-	JM-KD	disordered
2AUH	"	-	KD+adpt-prot	open ; yes x3
2B4S	"	ATP-pept	KD+phosphat	open ; yes x3
1JQH	Igf1r	ANP	KD	open ; yes x2
1K3A	"	ACP	KD	open ; yes x3
1M7N	"	-	KD	closed ; no
1P40	"	-	KD	closed ; no
1FGK	Fgfr1	-	KD	intermed,dis. ; no
1AGW	"	SU4984	KD	intermed ; no
1FGI	"	SU5402	KD	intermed ; no
2FGI	"	PD173074	KD	intermed,dis. ; no
1GJO	"	-	KD	open ; no
1OEC	"	AA2	KD	open ; no
2IVS	Ret	cAMP	JM-KD	open ; no
2IVT	"	AMP	JM-KD	open ; yes
2IVU	"	ZD6474	JM-KD	open ; yes
2IVV	"	PP1	JM-KD	open ; yes
1VR2	Vegfr1	-	KD	disordered ; no

Continued on next page . . .

PDB code	kinase	ATP-ligand	domains	A-loop
1Y6A	Vegfr2	AAZ	KD	disordered ; no
1Y6B	"	AAX	KD	disordered ; no
1YWN	"	pyrimidine-der	KD	close-disordered ; yes
2P2I	"	nicotinamide	JM-KD	disordered ; no
2P2H	"	triazine der	JM-KD	disordered ; no
1M14	Egfr	-	KD+40C-term	open ; no
1M17	"	Erlotinib	KD+40C-term ; open	no
1XKK	"	Lapatinib	KD+40C-term	intermed,dis. ; no
2GS2	"	-	JM-KD	open ; no
2GS6	"	ATP-pept	JM-KD	open ; no
4GS7	"	AMP-PNP	JM-KD	intermed ; no
2ITN	"	AMP-PNP	JM-KD(G719S)-Ctail	open ; no
2ITO	"	gefitinib	JM-KD(G719S)-Ctail	open ; no
2ITP	"	AEE788	JM-KD(G719S)-Ctail	open ; no
2ITQ	"	AFN941	JM-KD(G719S)-Ctail	open ; no
2ITT	"	AEE788	JM-KD(L858R)-Ctail	open ; no
2ITU	"	AFN941	JM-KD(L858R)-Ctail	open ; no
2ITV	"	AMP-PNP	JM-KD(L858R)-Ctail	open ; no
2ITW	"	AFN941	JM-KD-Ctail	open ; no
2ITX	"	AMP-PNP	JM-KD-Ctail	open ; no
2ITY	"	gefitinib	JM-KD-Ctail	open ; no
2ITZ	"	gefitinib	JM-KD(L858R)-Ctail	open ; no
2J6M	"	AEE788	JM-KD-Ctail	open ; no
2J5E	"	quinazoline-der	JM-KD-Ctail	open ; no
2J5F	"	quinazoline-der	JM-KD-Ctail	open ; no
1PKG	Kit	ADP	JM-KD	open ; no
1T45	"	STI 571	JM-KD	closed ; no
1T46	"	-	JM-KD	closed ; no
1RJB	Flt3	-	JM-KD	closed ; no
2I1M	Fms	arylamide-der	JM-KD	closed ; no
2I0Y	"	arylamide-der	JM-KD	closed ; no
2I0V	"	quinolone-der	JM-KD	closed ; no
2OGV	"	-	JM-KD	closed ; no
1R0P	Met	K-252A	JM-KD	closed ; no
1R1W	"	-	JM-KD	disordered ; no
2G15	"	-	JM-KD	intermed ; no
1JPA	EphrB2	adenine	JM-KD	disordered ; no
1MQB	EphrA2	ANP	JM	disordered ; no
2GSF	EphrB3	-	JM-KD	open-disorder ; no
2HEL	EphrA4	-	KD	open-disorder ; no
2HEN	EphrB2	ADP	KD (D754A)	open-disorder ; no

Continued on next page . . .

PDB code	kinase	ATP-ligand	domains	A-loop
1LUF	Musk	-	JM-KD	close-disordered ; no
1FVR	Tie2	-	KD	open ; no
2OO8	"	pyrid-pyrimid der	KD	interm-disorder ; no
2OSC	"	pyrid-pyrimid der	KD	interm-disorder ; no
2P4I	"	pyrid-pyrimid der	KD	open-disorder ; no
1YVJ	Jak3	Stauro-anal	KD	open ; yes x2
2B7A	Jak2	pyridone-der	KD	open ; yes x2
1U54	Ack1	ACP	KD	open, dis. ; yes
1U4D	"	DBQ	KD	open, dis. ; no
1U46	"	-	KD	open, dis. ; no
1FPU	Abl	PRC	KD	closed ; no
1IEP	"	STI 571	KD	closed ; no
1M52	"	PD173955	KD	open ; no
1OPJ	"	STI 571	KD	closed ; no
1OPK	"	PD166326	SH2-3,KD,reg-tail	open ; no
1OPL	"	PD166326	SH2-3,KD,reg-tail	open ; no
2F4J	"	VX-680	KD-C-term	open ; no
2FO0	"	PD166326	SH2-3,KD,reg-tail	open ; no
2G1T	"	ATP-pept	KD-C-term	intermed ; no
2G2H	"	PD166326	KD-C-term	open ; no
2G2F	"	ATP-pept	KD-C-term	open ; no
2G2I	"	ADP pept	KD-C-term	open ; no
2GQG	"	BMS-354825	KD	open ; yes
2HIW	"	benzamide der	KD	closed ; no
2HYY	"	STI 571	KD	closed ; no
2HZ0	"	NVP-AEG082	KD	open ; no
2HZ4	"	NVP-AFN941	KD	open ; no
2HZI	"	NVP-AFG210	KD	open ; no
2HZN	"	PD180970	KD	closed ; no
1BYG	Csk	Staurosporine	KD	intermed,dis. ; no
1K9A	"	-	SH2-3,KD,reg-tail	open-disorder ; no
1SNU	Itk	Staurosporine	KD	disordered ; no
1SM2	"	Staurosporine	KD	disordered ; yes
1SNX	"	-	KD	disordered ; no
1K2P	Btk	-	KD	open ; no
1XBA	Syk	-	KD	open ; no
1XBB	"	STI 571	KD	open ; no
1XBC	"	Staurosporine	KD	open ; no
1U59	Zap70	Staurosporine	KD	open ; no
1MP8	Fak	ADP	KD	disordered ; no
2ETM	Fak1	pyrimidine-der	KD	close-disordered ; no

PDB code	kinase	ATP-ligand	domains	A-loop
1FMK	Src	-	SH2-3,KD,reg-tail	intermed ; no
2PTK	"	-	SH2-3,KD,reg-tail	intermed ; no
2SRC	"	ADP-PNP	SH2-3,KD,reg-tail	intermed ; no
1KSW	"	ATP anal	SH2-3,KD,reg-tail	intermed ; no
1Y57	"	STI-anal	SH2-3,KD,reg-tail	open ; no
1YI6	"	-	KD-C-term	open ; yes
1YOJ	"	-	KD-C-term	disordered ; no
1YOL	"	CGP77675	KD-C-term	disordered ; no
1YOM	"	purvalanol A	KD-C-term	disordered ; no
2BDF	"	AP23451	KD	open-disorder ; no
2BDJ	"	AP23464	KD	open-disorder ; no
2H8H	"	AZD0530	SH2-3,KD,reg-tail	intermed ; no
2OIQ	"	STI 571	KD and Cterm	interm-disorder ; no
2HWO	"	quinazoline-der	KD(S345C); Cterm	disordered ; no
2HWP	"	quinazoline-der	KD(S345C); Cterm	disordered ; no
2DQ7	Fyn	Staurosporine	KD	open ; yes
3Lck	Lck	-	KD	open ; yes
1QPJ	"	Staurosporine	KD	open ; yes
1QPD	"	Staurosporine	KD	open ; yes
1QPC	"	ANP	KD	open ; yes
1QPE	"	PP2	KD	open ; yes
2OFU	"	pyrimidine-der	KD	open ; no
2OFV	"	pyrimidine-der	KD	interm-disorder ; no
2OG8	"	pyrimidine-der	KD	interm-disorder ; no
2OF2	"	furano-der	KD	open ; no
2OF4	"	furano-der	KD	open ; no
1AD5	Hck	ANP	SH2-3,KD,reg-tail	intermed,dis. ; no
2HCK	"	quercetin	SH2-3,KD,reg-tail	intermed,dis. ; no
1QCF	"	PP1	SH2-3,KD,reg-tail	intermed ; no
2C0I	"	A-420983	SH2-3,KD,reg-tail	disordered ; no
2C0O	"	A-770041	SH2-3,KD,reg-tail	disordered ; no
2C0T	"	A-641359	SH2-3,KD,reg-tail	disordered ; no
2HK5	"	isoxazole-der	KD	open ; no

Chapter 4

Assessing stability of wild type and oncogenic FLT3

4.1 Abstract

The way of activation for Flt3 oncogenic structures with presence of internal tandem duplication (ITD) is still an open question. In this study many efforts have been addressed to investigate the molecular mechanism of kinase domain activation and to depict which movements may be the first events.

Firstly, to validate the usage of MD simulation and to fine-tune a working protocol, several trajectories have been produced for insulin receptor in the non-productive and productive states. The reliability of the method was confirmed by comparing the evolution of the same coordinates with different initial velocities. The analysis of the radius of gyration, distance between the center of mass of the two lobes and the RMS deviations depicted certain characteristics of the productive and non-productive state of this protein. Later, the wild type protein of Flt3 was investigated and the simulations proved the stability of the KD strengthened by the presence of the juxtamembrane domain. Moreover, the stability of the isolated inhibitory domain was also tested revealing the maintenance of its main secondary structure.

The studies on oncogenic Flt3 started with hypothetical structures of Flt3 bearing ITDs, which were produced by means of homology modeling. Beside a steric hindrance of the C-helix, the analysis of the MD trajectories did not show any appreciable difference compared to the wild type. Further investigations were carried out on the calculated trajectories of different protein constructs with shorter JM. The release of KD from the auto inhibitory interactions revealed a conformational change whose major player is the C-helix.

4.2 Introduction

The gene mapped to the 13th chromosome at the locus 13q12 of the human genome, encodes for Flt3 protein tyrosine kinase [191]. This 993 amino acid long macromolecule has been classified as an element of the Pdgfr subfamily (also known as class III) of receptor tyrosine kinases (RTKs) [188]. Flt3 wild type activity has been reported for physiological proliferation, differentiation and apoptosis of hematopoietic cells [194] with

79

a major expression in lymphoid and myeloid progenitors [193]. Topologically, Flt3 is composed of several moieties: five extra-cellular immunoglobulin-like domains, that form the ligand-binding domain, a transmembrane domain and a cytosolic part with the juxtamembrane domain (juxtamembrane, JM), the catalytic domain (kinase domain, KD) and the C-terminal tail.

The life-cycle of Flt3 starts with the ribosomal expression of the full-length construct while the later maturation via N-terminal glycosylation localizes the protein at the cell membrane [195, 196]. The receptor can bind the endogenous ligand, a second protein called Flt3-ligand (FL) [199]. FL interacts with and co-localizes two Flt3 non-active units promoting the homo-dimerization of the receptor, which is a well known behavior for RTKs [146, 147]. The dimer formation increases the activity of the cytosolic kinase domains that transphosphorylates the proteins at different sites to stabilize the active form of the receptor [170]. The receptor transduces the signal inside the cell permitting the binding of second effectors to its phosphotyrosines and transferring phosphate groups to tyrosine residues of other proteins [130]. The cycle of the protein is finished by internalization and degradation of the complex Flt3-FL few minutes after the stimulation [202].

Insights into the mechanism maintaining the non-active state of the protein were revealed with the publication of the first, and so far unique, experimentally determined structure of Flt3 kinase domain [117]. The authors produced crystals where the auto-inhibition of the KD is mediated by the juxtamembrane, which is recognized as the prototype mechanism for Pdgfr subfamily [73, 117].

The scientific interests about Flt3 are the behaviors, characteristics and functions of the normal protein as well as the pathological implications of its deregulated activity. Aberrant Flt3 forms enhancing the activity of the receptor have been widely correlated with hematopoietic malignancies [208]. Two types of structural variations are found to cause diseases: i) point mutations [212, 213], within the KD, like Asp835Tyr, and ii) insertions of amino acid sequence identified as internal tandem duplication (ITDs) at the juxtamembrane [163]. The ITDs are often found in acute myeloid leukemia (AML) with a hit rate of approximately 30% of the cases [163]. The origin of ITDs is due to the aberrant biosynthesis of the normal Flt3 sequence; a stretch of nucleotides is transduced twice, one after the other [163]. The mechanism of the constitutive activation of the RTK by ITDs has been proposed as the disruption of the interaction between JM and the kinase domain thus interfering with the inhibitory role of this domain [73].

In this study, further characteristics of ITD activation of Flt3 were investigated with different protein constructs and by means of virtual simulation of the protein molecular motion via molecular dynamics protocols.

4.3   Materials and Methods

4.3.1   Homology modeling

The experimental derived structure of the wild-type Flt3 is available at the protein data bank (PDB) under the code: 1RJB. In the crystallographic structure the loop between β-strand 4 and the C-helix, and the whole kinase insert region are not defined and thus no atomic coordinates are available. The residues encompassing the first of the two missing parts were added with the module for loop building in the software package

SYBYL7.3 [249]. While the long tyrosine kinase insert region (KID), approximately 72 residues, was substituted with a stretch of four glycines in order to connect α-helices 5 and 6. The reasoning of this omission resides on the fact that no structural information is available about this amino acid stretch. In addition, experiments about the kinase domains of Kit, Vegfr and Flt3 were carried out with a recombinant form of the wild type lacking of their respective KIDs [73, 117, 151, 197, 250]. The resulting proteins showed normal behaviors and three-dimensional architectures.

The coordinates for the amino acids of this new loop were selected always with the module for loop building in SYBYL7.3 in such a way that the side chains previous and after the gap were not sterically influenced.

Protein structures of the Flt3 with ITD were virtually generated with comparative modeling by means of MODELLER9v1 [225]. Coordinates are assigned for the target sequence based on the structure template ones. The solution is a protein model whose conformation satisfy the spatial restraints derived from the template and assigned with the alignment of the two sequences.

The ITDs were added with the fast-refinement routine of the protein model. Only one structure, further submitted to the MD protocol, was chosen among ten different solutions based on the best stereochemical quality (bond length, ϕ and ψ optimal angles, etc.) assessed with PROCHECK [251].

4.3.2 Dynamic simulations

In the present work, molecular dynamics (MD) simulation protocols were used to study the motion of different protein structures. This methodology allows the user to explore the conformational space of proteins around specific energetic minima. The GROMACS-3.3.1 software package [252–254] was used to calculate and analyze trajectories.

The use of MD methodology to infer kinase domain dynamics was evaluated and validated by calculation and further analysis of trajectory for non-productive and productive states of insulin receptor (Insr) KD. Different systems composed by the protein atoms at a given conformation, a fix number of water molecules and counter-ions (Na^+ or Cl^-), to balance the total charge, have been simulated.

The coordinates of the protein atoms were taken from the crystallographic structures. The water molecules present in the file were also kept while in the case of the productive Insr (1IR3) the rest of the atoms was removed, namely the phosphates on the A-loop tyrosines and the two ligands (AMP-PNP and the peptide substrate).

Spc216 theoretical model [255] was used for the 12155 molecules simulating the explicit solvent around the macromolecule. Protein atoms and bonds were characterized with OPLS-AA force field parameters [256]. The non-bonded interactions were evaluated for electrostatics using Particle Mesh Ewald (PME) [257] with a coulomb radius of 9 Å and van der Waals within a distance of 14 Å. Berendsen algorithm [258] was employed to maintain the temperature and pressure constant during simulations with values of 300 K and 1 bar, respectively. All carbon-hydrogen bonds were constrained using the SHAKE algorithm [259]. Finally, all MD runs were performed, with a time-step of 2 femtoseconds.

The following protocol was performed for all the simulated proteins. Systems were first minimized, to avoid bad contacts of atoms, and then a MD run of 50 ps with position restrained for the protein atoms, to relax the solvent was performed. To start the MD

simulation for the trajectory production, initial velocities were assigned to the atoms with Maxwellian distribution a seed number for random generation. Systems were warmed up from 10 Kelvin to 300 during 7000 steps of Langevin dynamics with friction coefficient variable depending on the atom type (with a value of atom mass over 0.01 ps). A short MD simulation equilibration phase of 100 ps was performed with strong coupling for the thermostat and barostat and, eventually, the coupling constants were softened for the production simulations. The collected data from the simulation yielded to NPT ensemble of the systems.

Furthermore, the wild type as well as two other Flt3 constructs, namely Flt3-593 and Flt3-603 (see later for explanations), were simulated with NVT ensemble using only the thermostat and switching-off the barostat. Atomic coordinates (in the constructs part of the protein atoms were removed) as well as energies and velocities were taken from the previous run of the Flt3 wild type.

The idea behind is the exploitation of the parameters coming from an already relaxed system to avoid possible deviations due to the initialization and equilibration phases of MD. A position restraint MD for adjusting position of water molecules, in the case of the two constructs, and a short (7 ps) Langevin dynamics for all proteins to permit the slow release of the atoms were employed before the final simulations.

The values for different parameters were extracted and analyzed with different programs of the GROMACS-3.3.1 software package. For visual representation and inspection of structures and trajectories the computational programs VMD-1.8.7 [260] and BODIL-0.8.1 [61] were used. Further data sets elaboration and display were carried out with XM-GRAGE provided as software package of Debian GNU/Linux distribution. In the present chapter, atomic coordinates in the figures were rendered with PYMOL [64].

4.3.3   Superpositioning

The selection of the atoms to superpose is guided by the observation of the crystallographic structures of the several kinase domain proteins in different conformations. In general, conformational changes for this type of structure are appreciated as the relative movements of one lobe respect to the other one. At the C-lobe, beside the variations of the A-loop, often dynamically isolated from the rest, no other major change is appreciated for secondary structures independently from the main core of the lobe (α3 and α5 for Insr (PDB code: 1IRK), and α6 and α9 in Flt3 (1RJB)). At the N-lobe variations of the C-helix and Gly-rich loop are usually seen in respect to the main core of the subdomain: the main β-sheet.

4.4   Results and Discussions

4.4.1   Insulin receptor motion

MD protocol evaluation

Several MD simulations were performed on the non-productive and productive conformations of insulin receptor KD, whose molecular structures have been determined experimentally (PDB code 1IRK and 1IR3, respectively), with the following aims:

    - to assess the reliability of the molecular dynamic (MD) simulation protocol for studying the molecular motion of the catalytic domain of Flt3;

    - to produce a comparison system to properly judge the further dynamics;

After initial trials on the atomic coordinates of the two Insr structures to optimize parameters and settings of the protocol, 8 ns MD simulations were performed to evaluate the general plasticity of the two systems.

The evolution of several indices during calculations were monitored to validate the performed experiments.

In figure 4.1, the kinetic, potential and total energy are plotted for the first 200 picoseconds of simulation of the non-productive state of Insr. The reported energetic components
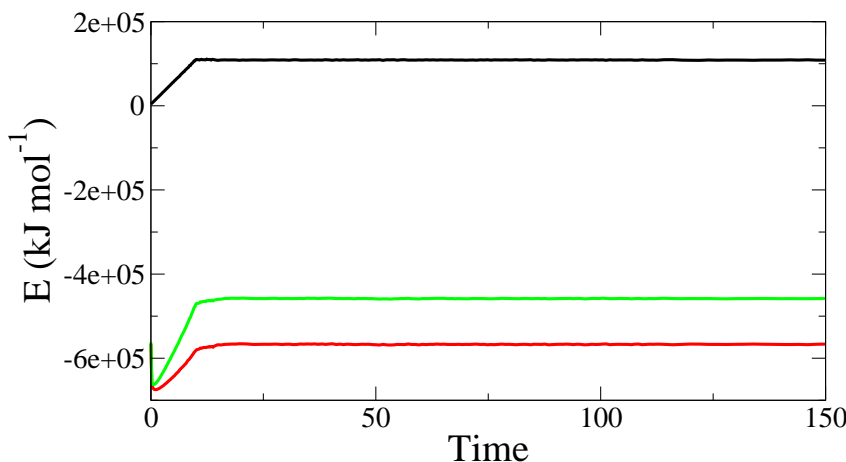
Figure 4.1: *Energies plot for the Insr. The kinetic energy (in black), potential energy (in red) and total energy (in green) are reported as function of time in ps for the first steps of the MD simulation of the non-productive state of Insr.*

encompass all steps of the used protocol, the heat-up (0-14 ps), equilibration (14-114 ps) and the beginning of the production phase (114-8000 ps). The potential and total energy of the system decrease for the first picoseconds of calculations. All energy terms raise up with the increasing temperature till a plateau which lasts for the equilibration and production phases. Temperature and pressure of the system were also monitored and found to be stable at the given values (data not shown). These parameters were checked for all the MD simulations carried out in these studies and found in agreement with the ones just reported.

A second evaluation of the used protocol and its application to the studied system is the assessment of the structural stability and the time required by the system to equilibrate. This characteristics were judged based on the profile of the root mean square (RMS) deviation calculated as function of time for the positions of the Cα-atoms belonging to the kinase domain. To note, the removal of the phosphate groups from the three A-loop tyrosine residues of the productive conformation, present in the crystallographic structure, did not prevent the system for reaching a local equilibrium and to maintain it for the full length of the simulation. The principal reason why the whole movement is not reproduced in these simulations is most likely due to the short time scale investigated compared to the required period for such a conformational change to happen, estimated

for Pka kinase within milliseconds range [93]. Actually, this removal would promote a big conformational change at the A-loop leading to inactivation of the enzyme. In addition, it is also postulated that an initial unphosphorylated productive form exists in order to trigger the first event of transphosphorylation [70].

As shown in figure 4.2, the Insr structures reached a rough equilibrium before the second nanosecond of simulation (around 1.6 Å for the productive state and 1.8 Å for the non-productive one) and maintained it with fluctuations around a mean value. Because of the early convergence of the systems to their equilibria, for the successive simulations of the Insr KD data were collected for 3 nanoseconds.
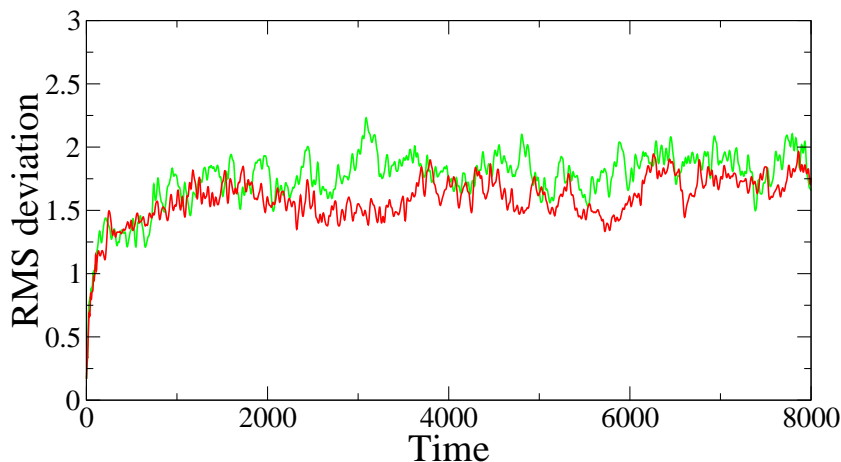
Figure 4.2: *Structural stability assessment from 8 ns long simulations on Insr KDs. Plots of RMS deviations of Cα-atoms from the initial configuration is reported in Å as function of time (in ps) for the non-productive, in green, and for the productive, in red, states.*

MD protocol reproducibility

The second issue addressed by these pilot simulations was the reproducibility of data production with the used methodology. Due to the chaotic nature of molecular motions in general and in particular in MD simulations, where the force acting on one atom is calculated as the sum of the forces of the surroundings, the evolution of similar initial atomic configurations can diverge during different computations, later confirmed. Thus it is important to check to what extent the resulting trajectories and their differences depend from the assigned initial velocities (artefact) or are the expression of relevant biological differences.

The key, to check the variability of the results, is the random assignment of the initial velocities to the atoms of the system. The seed number used to generate the velocities was changed resulting in different distributions of these values. Three different initial distributions of velocities for the non-productive and productive form were employed as well as the same distribution (only for the non-productive state) but in different computers with the same and different hardware architectures. The trajectories calculated using different machines with the same architecture result in the overlapping of the parameters (data not shown). While, computers with different architectures compute different trajectories starting with the same atomic coordinates and velocities (see figure 4.3a on page 86).
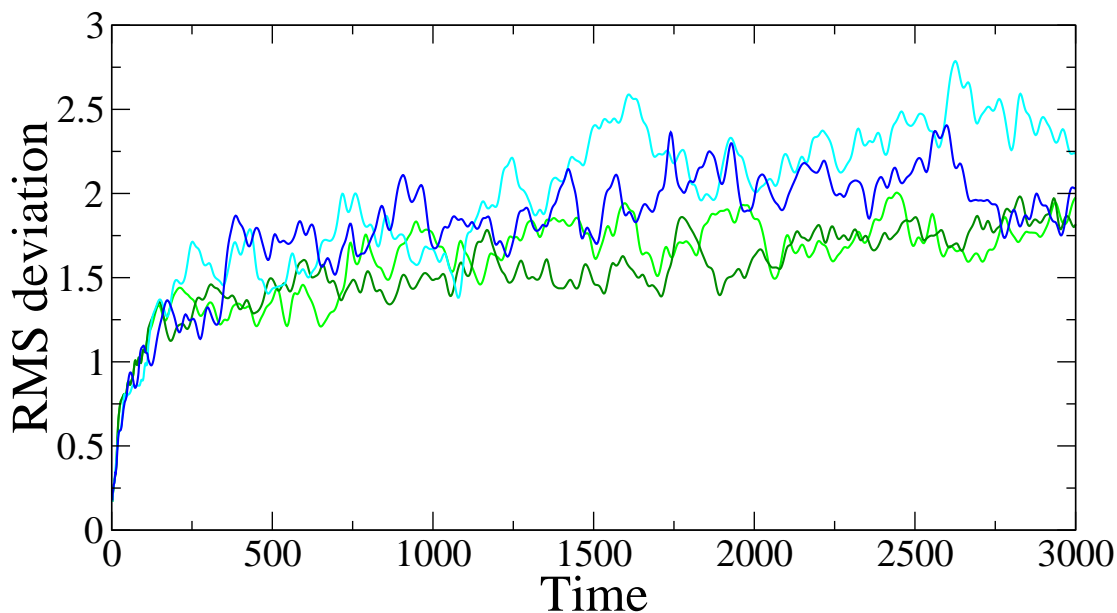
(a) *Insr non-productive state*
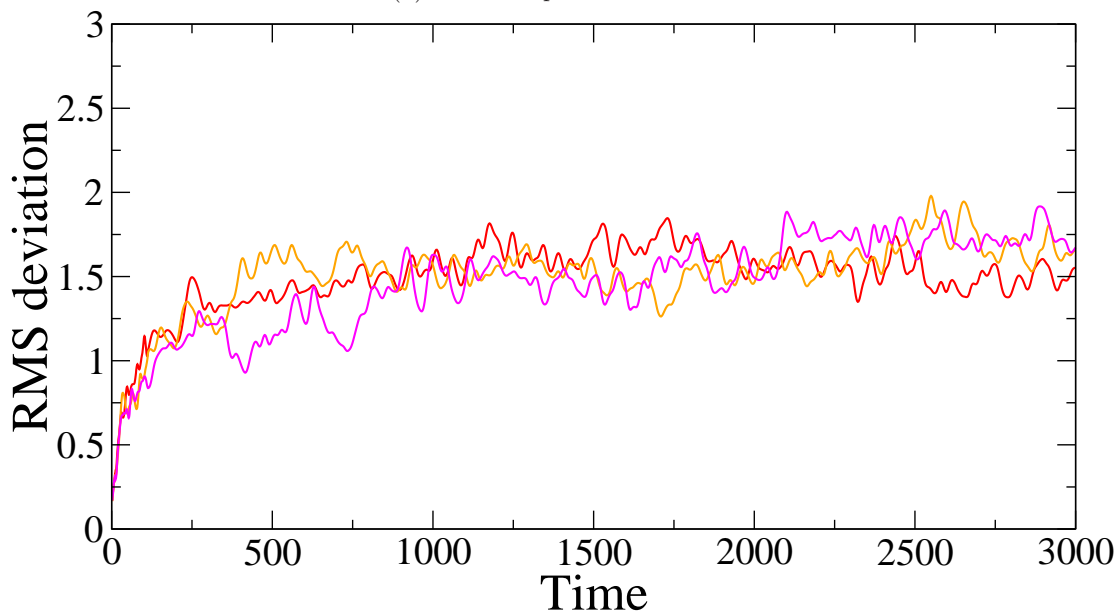
(b) *Insr productive state*

Figure 4.3: *Assessing the MD reproducibility. Plots of RMS deviations of Cα-atoms of the catalytic domain is reported in Å as function of time (in ps). Progression of RMSD are drawn for the non-productive (with different seed numbers in green, cyan and blue and different computer architecture in dark green) and for the productive (with different seed numbers in red, magenta and orange) states.*

Different velocities for the same system affected the non-productive state to a larger extent compare to the productive as indicated by the slightly higher relative standard deviation of the root mean square (RMS) deviation values. The RMS deviations reached an average value, for the last nanosecond, of 2.07 Å with a standard deviation (sd) of ± 0.275 (relative sd $\pm 13.2\%$) for the non-productive and 1.69 Å ± 0.101 (relative sd $\pm 5.9\%$) for the productive states. In general, the trajectories were considered in agreement in reproducing a similar behavior for the two protein states. The monitored indices of the MD clearly indicate that equilibrium states can be reached in agreement with the fact that an X-ray structure results from an average of similar conformations which are energetically stable over time: the thermodynamic equilibrium. Thus, the protocol can be judged reliable to study the molecular motion of kinase domains around a given energy minimum.

Sterical considerations

The sterical progression of the two protein conformations in time was evaluated monitoring the radius of gyration and the distance between the centers of mass of the N-lobe and the C-lobe during the simulations (figure 4.4 on page 87). The plots reveal stable trends for the observed features of the two states and with different seed numbers. The mean value of the radii for the non-productive state simulations is 19.13 Å ± 0.156 (relative sd $\pm 0.8\%$) while a lower value is achieved for the productive state, 18.89 Å ± 0.021 (relative sd $\pm 0.1\%$). Considering the inter-lobe distance, mean values of 25.76 Å ± 0.590 (relative sd $\pm 2.3\%$) for the non-productive and 24.10 Å ± 0.138 (relative sd $\pm 0.6\%$) for the productive state are calculated.

The non-productive conformation of Insr KD has a slightly bigger hydrodynamic volume with the two lobes further away compared to the productive conformation. Thus, a more compact tertiary structure is appreciate in the case of productive protein in comparison with the non-productive.

Concluding remarks on MD protocol

In summary, a protocol was tuned to simulate tyrosine kinase domain using non-productive and productive conformations of Insr. Production and analysis of the trajectories depict their general stability over the computed time. Equilibria were reached before the first nanosecond of simulation with a larger deviation from the initial conformation in the case of the non-productive state (approximately 2 Å) compared to the productive state (approximately 1.7 Å).

Another feature extrapolated from the trajectories is the compactness of the two protein conformations. The comparison of radius of gyration and inter-lobe distance profiles reveal the productive protein with a more packed arrangement of the secondary structures in respect to the non-productive one.

Systems and used MD protocol were also tested for reproducibility using different seed numbers for random generation of the initial velocities. Three different trajectories were simulated for each state resulting in smaller range of parameters values for the productive than for the non-productive. Thus, to consider the RMSD significantly different the deviation has to be at larger than 13% for the non-productive and larger than 6% for the productive.

(a) *Insr radius of gyration*   (b) *Insr inter-lobe distances*
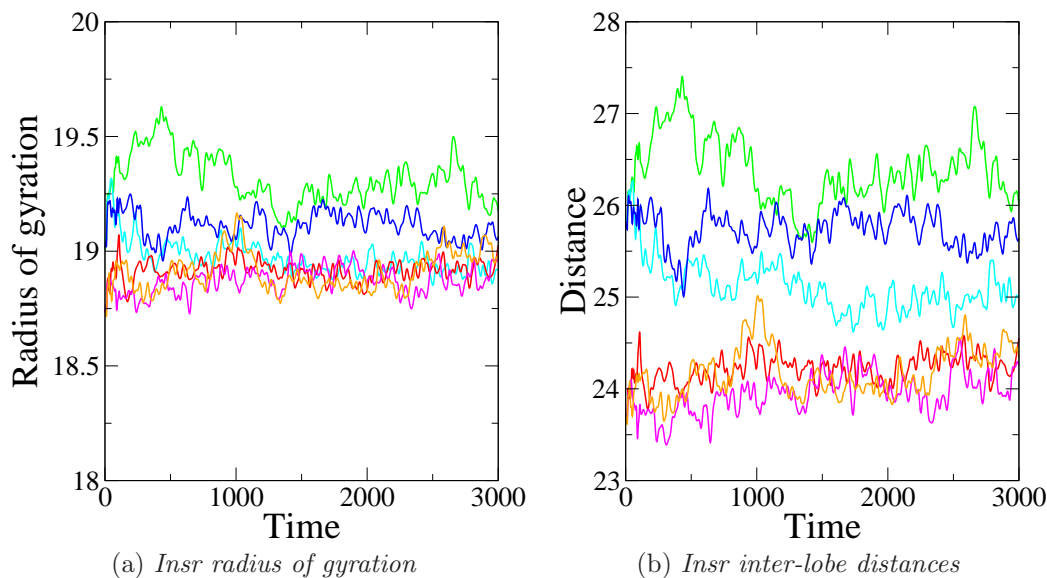
Figure 4.4: *Development of the radius of gyration and inter-lobe distance for KD of Insr during simulations. The progressions of the radius of gyration (a) and distance between the centers of mass of the N-lobe and C-lobe (b), are reported in Å and plotted as function of time, in ps. Profiles are drawn for the non-productive (with different seed numbers in green, cyan and blue) and for the productive (with different seed numbers in red, magenta and orange) states.*

4.4.2   Flt3 protein

The Flt3 protein is a tyrosine kinase receptor whose role is to transduce the signal inside the cell via phosphorylation by its kinase domain. The so-called juxtamembrane (JM) is a small N-terminal domain whose function is to maintain the KD inhibited when the receptor is not active and to bind secondary protein, via its phosphotyrosines, when the receptor is active.

The presence of ITDs has been correlated with alteration in the Flt3 signaling and with hematopoietic diseases like AML. Since 2004, the inhibited form of the protein is known: the kinase domain forms a tight complex with the JM [73]. Based on the structural information it has been postulated that the oncogenic activation by ITDs results from the disruption of the interaction between part of the juxtamembrane and the KD [73]. Nevertheless, further confirmations of such a mechanism at the molecular level have not yet come. To shed light on the molecular mechanism behind the KD activation by ITD insertion, comparative molecular dynamics between wild-type Flt3 and virtually generated Flt3 models were carried out.

Structural considerations for wild-type KD and JM

The typical features of the protein kinases are here depicted for the Flt3 amino acid sequence [52]:
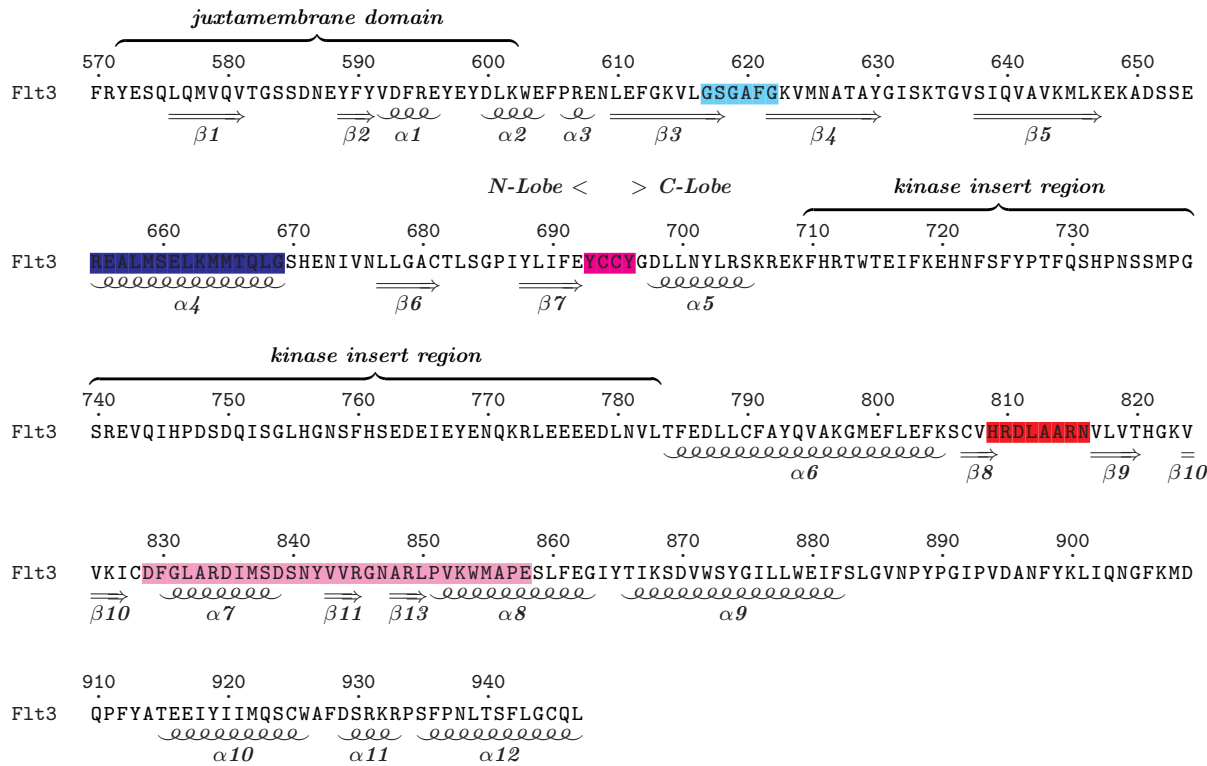
Figure 4.5: *Flt3 amino acid sequence. The primary structure (taken from* EXPASY *database under the entry P36888) of Flt3 protein is reported for the portion spanning the juxtamembrane domain and the kinase domain. Residues are identified by single-letter code and some regions are colored, in agreement with figure 4.6 on the next page, cyan, the Gly-rich loop, blue, the C-helix, magenta, the hinge region, red, the catalytic loop, and pink, the A-loop. Secondary structure assignment are reported at the bottom according to the crystallographic structure (1RJB). At the top boundaries of N- and C-lobe, juxtamembrane and the kinase insert region are indicated.*

As discussed in chapter 1 on page 1, this conserved stretches of residues are implicated in different functions such as, ligand binding, structural characteristic and catalysis.

The confirmation of the folding of the catalytic moiety came with the publication of the crystallographic structure [73] whose inhibited form is maintained by the juxtamembrane domain (depicted in figure 4.6 on the next page). While the catalytic domain is conserved in its fold compared to other RTKs [73], here we will focus on the structural information of the JM-KD interactions derived from the solved X-ray structure.

In Flt3 the sequence of the juxtamembrane domain starts at Tyr573 and ends at Lys602. It is found at N-terminal position of the N-lobe and shares sequence similarity with the other members of the Pdgfr subfamily of RTKs [73].

As depicted in figure 4.6 on the facing page the N-terminus of JM is positioned between the N- and C-lobe and interacts with the C-helix and the catalytic loop which are essential

for catalysis. The first secondary structure element of this domain is a strand, $\beta 1$, which
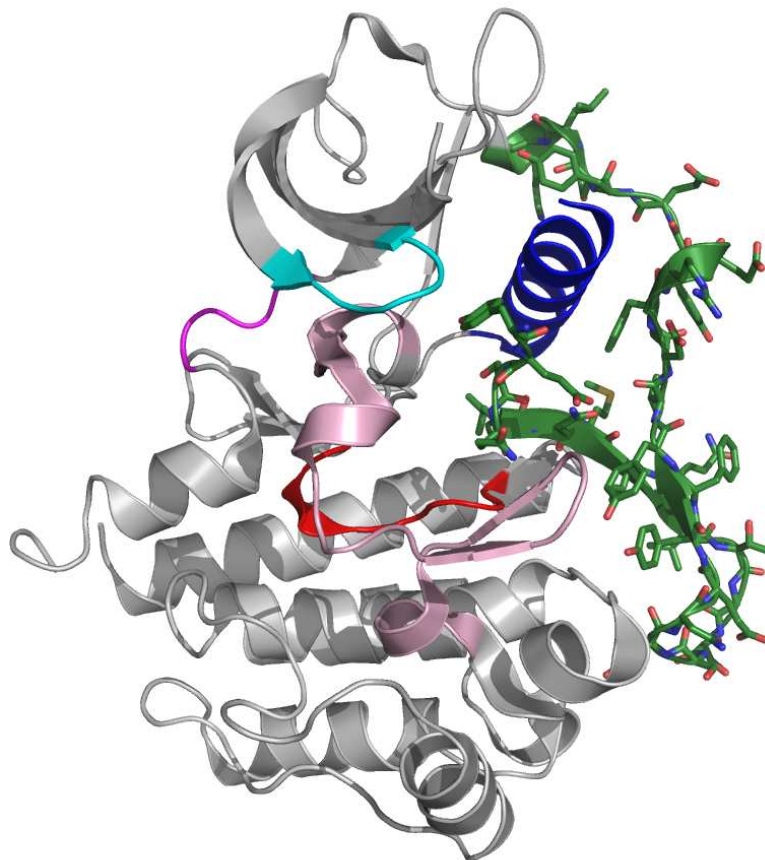
Figure 4.6: *Spatial arrangement of the juxtamembrane domain in complex with the kinase domain. The secondary structure of the protein is represented as cartoon and color-coded in gray for the KD. Other protein portions are also colored in green the juxtamembrane, cyan the Gly-rich loop, blue the C-helix, in red the catalytic loop and in pink the A-loop and magenta the hinge region. Furthermore, the residues of the JM and N-terminal of the KD are in sticks representation.*

forms a β-sheet with the strand before the C-loop, $\beta 8$, and with the second shorter strand of the JM, $\beta 2$ (see in figure 4.6). At the level of $\beta 2$ a pair of tyrosine residues is found and the first of the two, conserved within the Pdgfr subfamily, promotes the kinase activation when phosphorylated [250]. The rest of the JM has mainly coil structure, beside two distinct turns of α-helix, and interacts with amino acids of the outer face of the C-helix. Although some polar bonds are seen in the JM-KD complex, with residues at the N-terminus and in the loop between the two strands, the major interaction component is the extensive van der Waals contact of aromatic and aliphatic side chains of JM and C-helix (see figure 4.6).

The isolated juxtamembrane domain

An interesting study about a synthetic peptide encompassing the JM sequence of Kit revealed several characteristics for this domain [246]. This peptide in solution was folded and the secondary structure could be disrupted by chemical denaturation and tyrosine phosphorylation. Furthermore, the authors demonstrated the inhibition of the soluble catalytic domain of Kit mediated by this peptide.

The outcomes of this study depict the JM of Kit as a stand-alone domain with proper folding and full functionality. Based on the sequence comparison of this element throughout the Pdgfr subfamily it can be hypothesized that similar behavior can be expected for the JMs of the other family members.

In addition, these evidences suggest that a thermodynamic equilibrium between bound and unbound forms of the JM can exist [73]. In the basal state the bound form would be more stable, with the consequent inhibition behavior, while phosphorylation is likely to stabilize the unbound form.

Prompted by these hypotheses the stability of the isolated juxtamembrane domain of Flt3 was tested via MD simulation. The atomic coordinates of protein portion corresponding to the JM were taken from the crystallographic structure, surrounded by water molecules and trajectories were calculated for 3-5 ns. In agreement with the results for Kit, the performed simulations confirmed the stability of the JM of Flt3 over the calculated period of time as deduced from visual representation of the trajectory.

(a) *0 ns*   (b) *3 ns*
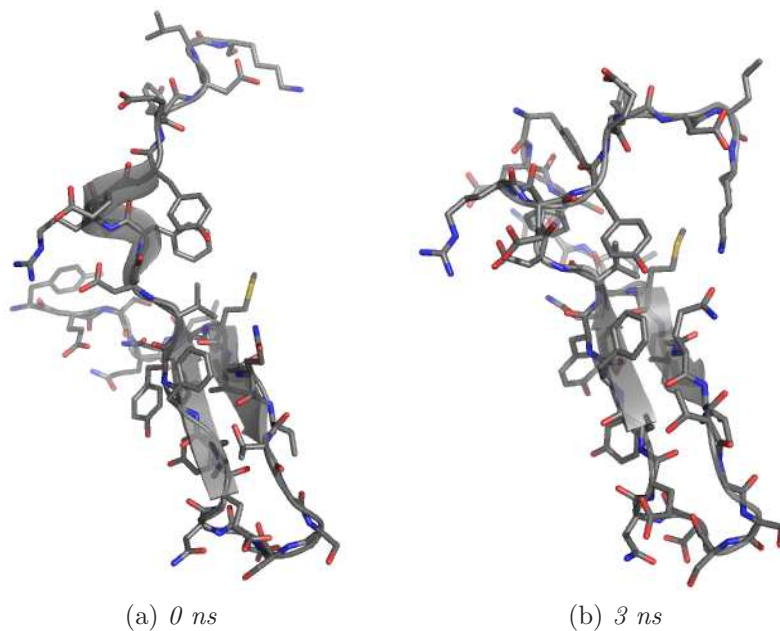
Figure 4.7: *Stability of the isolated juxtamembrane domain. The input structure and output structures of a representative MD simulation of the JM are drawn for secondary structures, mainly β-strand and coil contents. The atoms of residues are also represented as sticks.*

In figure 4.7 the starting and the final conformations of a representative simulation are depicted. In details, the two strands involved in the β-sheet formation are quite conserved

during simulations while the α-helix turn is not. The rest of the domain, which has coil secondary structure, interchanges different random conformations.

The results from these MD simulations arise the possibility that part of the JM could still be structured even in the presence of the ITD sequence when the JM is supposed to be unbound.

Oncogenic Flt3 proteins

The opportunity to investigate oncogenic forms of Flt3 with internal tandem duplication is due to a collaboration with clinicians. M.D. Emanuela Frascella and her group screened and identified a large set of ITDs from pediatric patients with promielocytic leukemia disease. Part of the results are reported here in table 4.1.

Table 4.1: *Internal tandem duplication sequences. In the left column the identifier for the sequences. In each case the transcription of the flt3 gene is normal till the amino acid whose number is reported as "AA stop" (right column) the new insertion starts from an earlier residue, reported as "AA start" (middle column) and progress with the normal transcription till the stop-codon.*

ITDs	AA start	AA stop
m012	582	595
m033	595	600
m089	598	614
m172	580	613
m237	594	610
m269	590	600
m301	585	594
m330	585	596
m355	599	609
m367	594	611
m383	591	602
m400	594	599
m410	580	598
m444	575	613

The ITDs can be found at several positions of the JM sequence and are even beyond its boundaries (residues 573-602) in the direction of the kinase domain. The length of such insertions is also variable from shorter ones, 5 amino acids, to longer ones with 30 and more residues. Because of the wide range of possible configurations for oncogenic Flt3 proteins here we focus on the subset occurring in the proximity of the coiled structure of the autoinhibitory domain.

The first issue is the generation of the possible conformation for proteins with such insertions. The ITD insertions are hypothesized to misalign the interaction partners of the KD-JM complex resulting in the relief of part of the auto-inhibitory domain. Mainly two scenarios were judged to be achievable. The first, where the main secondary structure of the JM, the β-sheet formation, is intact and a partial interaction is allowed for the rest

of the juxtamembrane. This is a feasible panorama in the case of short ITDs where just the thermodynamic equilibrium between bound and unbound form of the autoinhibitory domain is more affected towards the unbound form than in the physiological conditions. In this case, although the directionality of the DNA transcription places the inserted sequence at C-terminal position respect to the original sequence, structurally, it is not evident which of the two repeated stretch is going to interact with the predestine partner and conformation. The second scenario is more in agreement with the presence of long ITD sequences. Although part of the JM might still be properly folded, the arrangement of the elongated peptide for interaction is unlikely because of the energetic costs for the system. In this case the autoinhibitory domain would be more stable in the unbound form.

To investigate the first scenario, short sequences were selected from the clinical findings for structural and dynamic investigations. The sequences, namely m301, m355 and m400, were chosen to explore the effect of ITD insertion at different positions of the juxtamembrane structure. In the following alignment the positions of the ITD insertions is clarified at the JM while the invariant parts of the protein are omitted.

```
                  580          590                      600              610       620
                   .            .                        .                .         .
wild type    YESQLQMVQVTGSSDNEYFYVDF..........REYEY......DLKWEFPREN...........LEFGKVLGSGA
ITD m301     YESQLQMVQVTGSSDNEYFYVDFSDNEYFYVDFREYEY......DLKWEFPREN...........LEFGKVLGSGA
ITD m355     YESQLQMVQVTGSSDNEYFYVDF..........REYEY......DLKWEFPRENYDLKWEFPRENLEFGKVLGSGA
ITD m400     YESQLQMVQVTGSSDNEYFYVDF..........REYEYFREYEYDLKWEFPREN...........LEFGKVLGSGA
```

Figure 4.8: *Sequences of Flt3 oncogenic proteins. The alignment of the primary structure of Flt3 wild type and with ITDs is reported for the portion spanning the juxtamembrane domain and the N-terminus of the kinase domain. Residues are identified by single-letter code and ITD regions are colored in red.*

Thus, homology modeling was applied using the crystallographic structure of Flt3 (1RJB) as template and based on the alignment between oncogenic and wild type sequences (in order to generate the coordinates for the ITD sequences. More than one solution per constructs was selected for MD analysis to test the influence of different conformation of the newly added sequences. Furthermore, the repeated portions were added in C-terminal position at the original stretches, in line with the progression of the aberrant gene transcription, and in N-terminal position because of the possible alternative interaction envisaged before.

About the molecular dynamics simulation, different protocols were tried out with different settings, like higher temperatures than 300 Kelvins to furnish energy for enlarging the variety of possible conformations experienced by the system.

Here, trajectories are reported for the three ITD models and the wild type protein obtained with the application of the protocol described for Insr in materials and methods for 5 nanosecods.

The wild type JM-KD complex was simulated to acquire the dynamical behavior of the normal Flt3 protein and the basis for judging the ones of the generated conformations. From the visual inspection of the trajectory and the calculation of the RMS deviation for Cα-atoms positions of the KD (reported in figure 4.9 on the next page), the structure

was judged stable. Based on the RMSD profile, depicted in figure 4.9, the conformation present in the crystallographic structure is very close to a thermodynamic equilibrium. In fact, the dynamical evolution of the system quickly reaches a rough stability going to a value around 1.1 Angstroms for the last two nanoseconds of simulation. Comparing with the Insr dynamics, this system deviates to a lesser degree for finding its stability. This is likely to be explained by the presence of the JM which indeed restrain the catalytic domain motion.
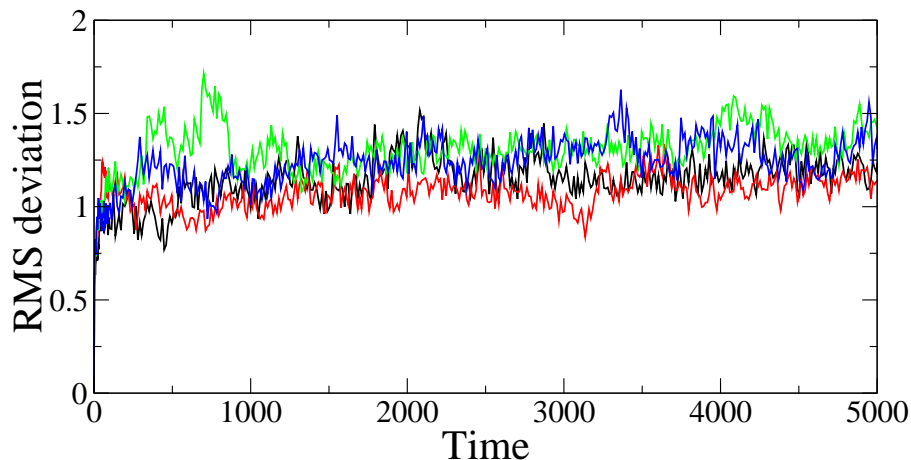
Figure 4.9: *The RMSD of Cα-atoms of the catalytic domains for wild-type and the three ITD constructs. The progressions of the RMSD in Å are plotted against time in ps for the wild-type in black, ITD m301 in red, ITD m355 in green and ITD m400 in blue.*

About the trials on the oncogenic forms, in figure 4.10 the conformations of the homology models after minimization are reported as well as the ones resulting by the application of the MD protocol. The RMS deviation from the initial coordinates displays the achievement of the stability at different degrees compared to the wild type structure (see figure 4.9). For the last 2 ns of the simulations the oncogenic systems have a RMSD mean value in a range of $\pm 12\%$ around the one of the wild type, which is in between the reproducibility of the Insr non-productive and productive states, afore mentioned. In the case of m355 a peak is seen due to the particular position where the ITD insertion occurs, which affects the β-sheet of the N-lobe. However, the visualization of the trajectories did not point out any common trend for the oncogenic forms which could remarkably differentiate them from the dynamical behavior of the wild type. In general, just fluctuations of the newly added loops were appreciated from the trajectory visualizations without any affection of the normal motion of the kinase domain.

The only common trend from the trials with different positions and conformations of ITDs in the generated structures was appreciated when the interactions between JM and the C-helix were partially disrupted, as in the case of m400 in figure 4.10 on the next page. This helix reacts to the missing interactions with movements in the directions of the ITD insertion and the close β-sheet follows this dynamics. This was interpreted as a steric hindrance of the JM on the secondary structure element, which, when released shows a own dynamic behavior.

(a) *ITD m301*

(b) *ITD m355*

(c) *ITD m400*
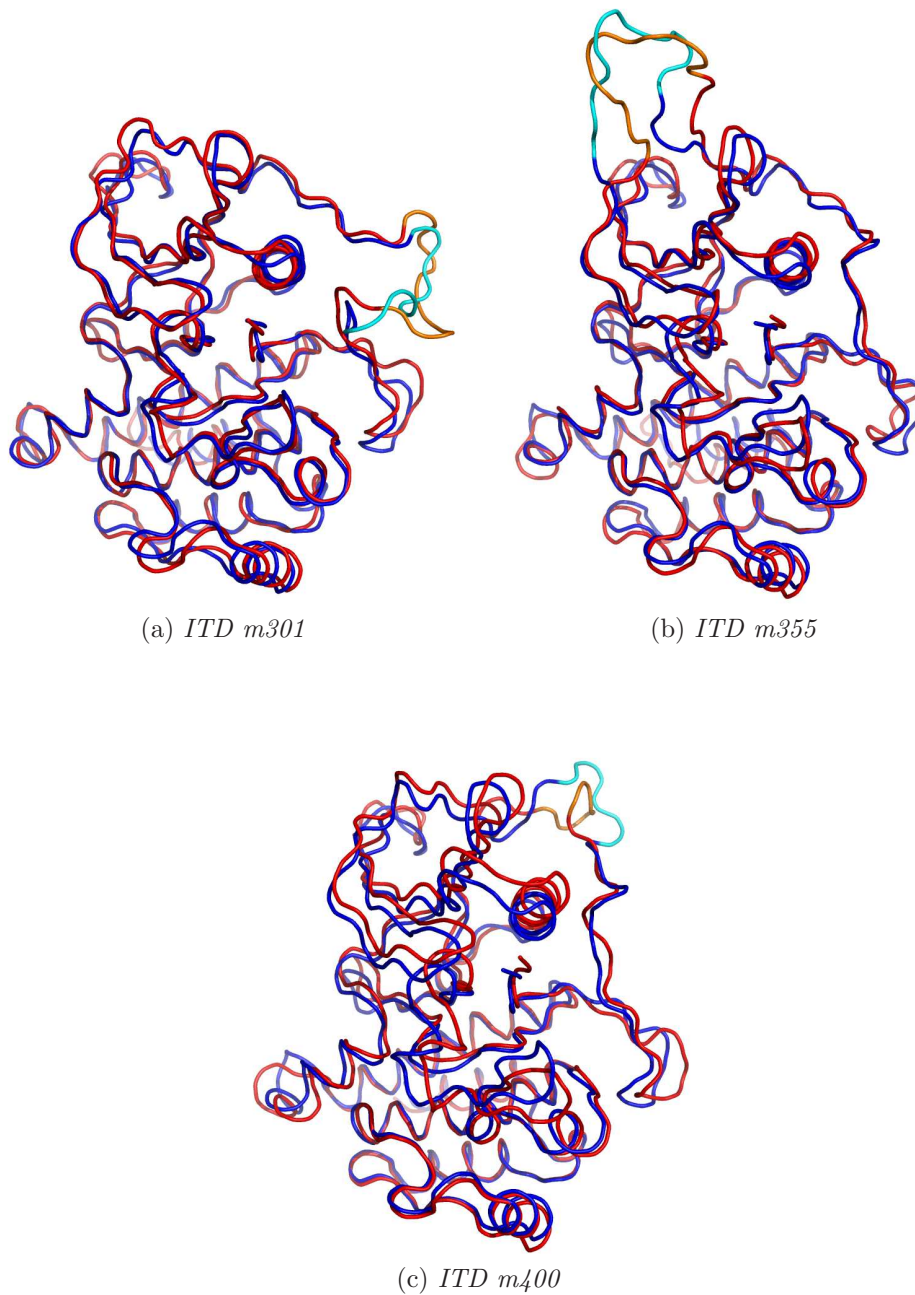
Figure 4.10: *Structures of the catalytic domain of oncogenic Flt3 at different time points of the MD simulations. The secondary structures are in ribbon representations. The colors identify the time point, in blue at 0 ns and in red at 5 ns. ITD insertions are also shown in cyan at 0 ns and orange at 5 ns.*

The result of these trials is the hypothesis of the C-helix reacting to the loss of partial interaction with the auto-inhibitory domain. Thus, further experiments to address this open issue were planned and carried out for different JM-KD complexes of Flt3 with variable length of the autoinhibitory domain.

Different JM-KD complex constructs

To investigate the hypothesis, the thermostability of the wild-type protein was compared with the KD in complex with a shorter JM domain and without. Furthermore, the physiological release of the juxtamembrane domain prior tyrosine phosphorylation should result in an activated KD conformation as seen for Kit (1PKG) [250]. Thus, the shortening of the JM may also be figured out as the attempt to mimic the situation in which the protein is activated. The JM leaves its position between the N- and C-terminal lobes in accordance to the movement observed in structure of the activated KD of kit also seen for Kit/Imatinib complex (1T46) [73].

In the crystallographic structure, 1PKG, just the C-terminal end of the JM interacts with the KD via C-helix while the rest of the domain (the visible part) is unstructured when the phosphotyrosine is present. The second KD molecule within the same asymmetric unit would even suggest that the JM may not contact the KD anymore after activation [250].

Three constructs have been investigated, namely wild type and two truncated forms of the proteins. The latter ones were obtained as follows. The N-terminal portion of the juxtamembrane was deleted till residue 593 in one case and 603 in the other (see figure 4.11 on the following page). The first construct generated contains the part of the JM covering the C-helix but lacks the β-sheet and the buried region and is identified as Flt3-593. While, for the second one the whole JM was removed, this is called Flt3-603. The three systems investigated were derived from the previous simulation of the Flt3 wild type. The atomic coordinates, atomic velocities and energies were taken from the end point of the 5 ns wild-type simulation. The type of the performed ensemble for this round of simulations was chosen to be NVT; no pressure adjustments were allowed in order to dissipate as less internal energy as possible. The trajectories for the three systems were thus calculated for 10 ns period of protein lifetime.

The first analysis of the simulated systems was the visual representation of the trajectories in VMD, a computational tool which allows the appreciation of the evolving coordinates as a movie. The stability of the wild type protein over time is appreciated as well as certain conformational changes in the other two constructs. Flt3-603 protein responds immediately (before the first nanosecond of simulation) to the removal of the whole juxtamembrane with a large change mainly located at the N-lobe. While the partial deletion of the JM in Flt3-593 leads to a similar structural variation but later in time (after the first nanosecond). From figure 4.12 on the next page the entity of such movements is revealed with the superposition of two protein conformations at different time points.

To numerically quantify those structural variations the deviation of the RMS for Cα-atoms of the KD over time and their RMS fluctuations were computed for the three trajectories (reported in figure 4.13 on page 97 and figure 4.14 on page 98).

In line with the visual inspections, the differential analysis of the RMSD shows that there are regions that experience more movements than others while the general folding within the three systems remains conserved. This is suggested by the fact that the RMSD

(a) *Flt3-wt*   (b) *Flt3-593*   (c) *Flt3-603*
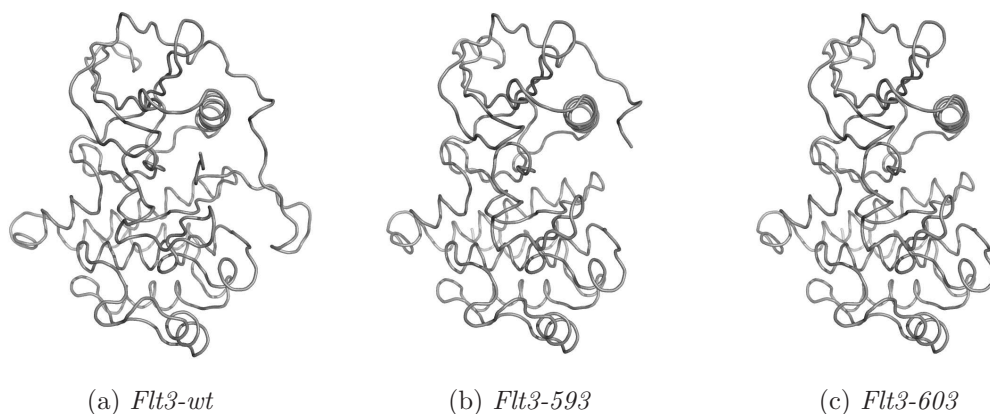
Figure 4.11: *Flt3 protein constructs. The three configurations are depicted in ribbon representation and colored gray.*

(a) *Flt3-593 at 1 ns and 5 ns*   (b) *Flt3-603 at 0 ns and 1 ns*
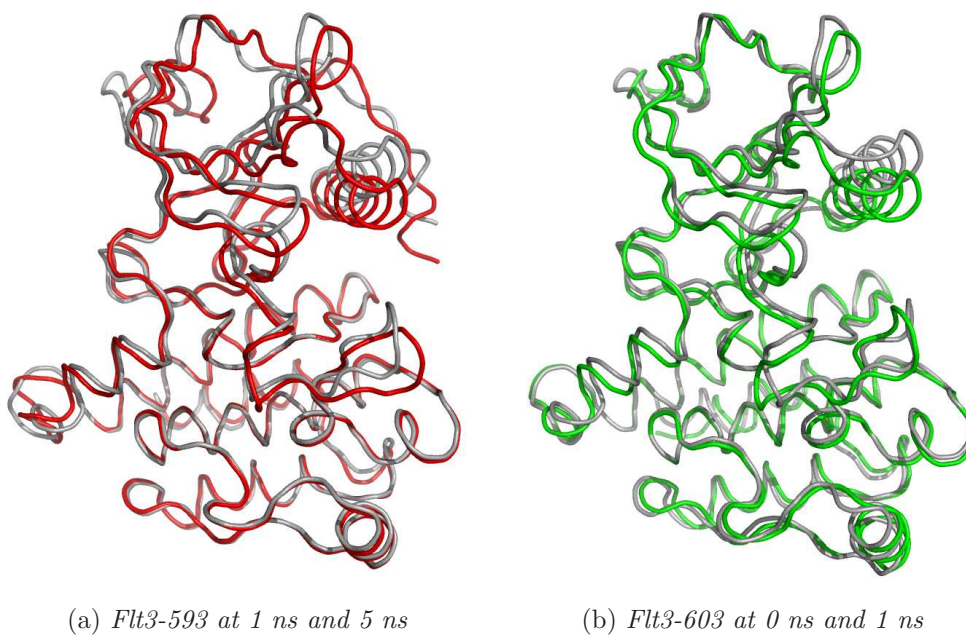
Figure 4.12: *Conformational changes of Flt3-593 and Flt3-603. Secondary structures of the proteins are shown with ribbon represent. The conformations are color-coded for time: Flt3-593 is in gray at 1 ns and red at 5 ns, and Flt3-603 is gray at 0 ns and green at 1 ns.*
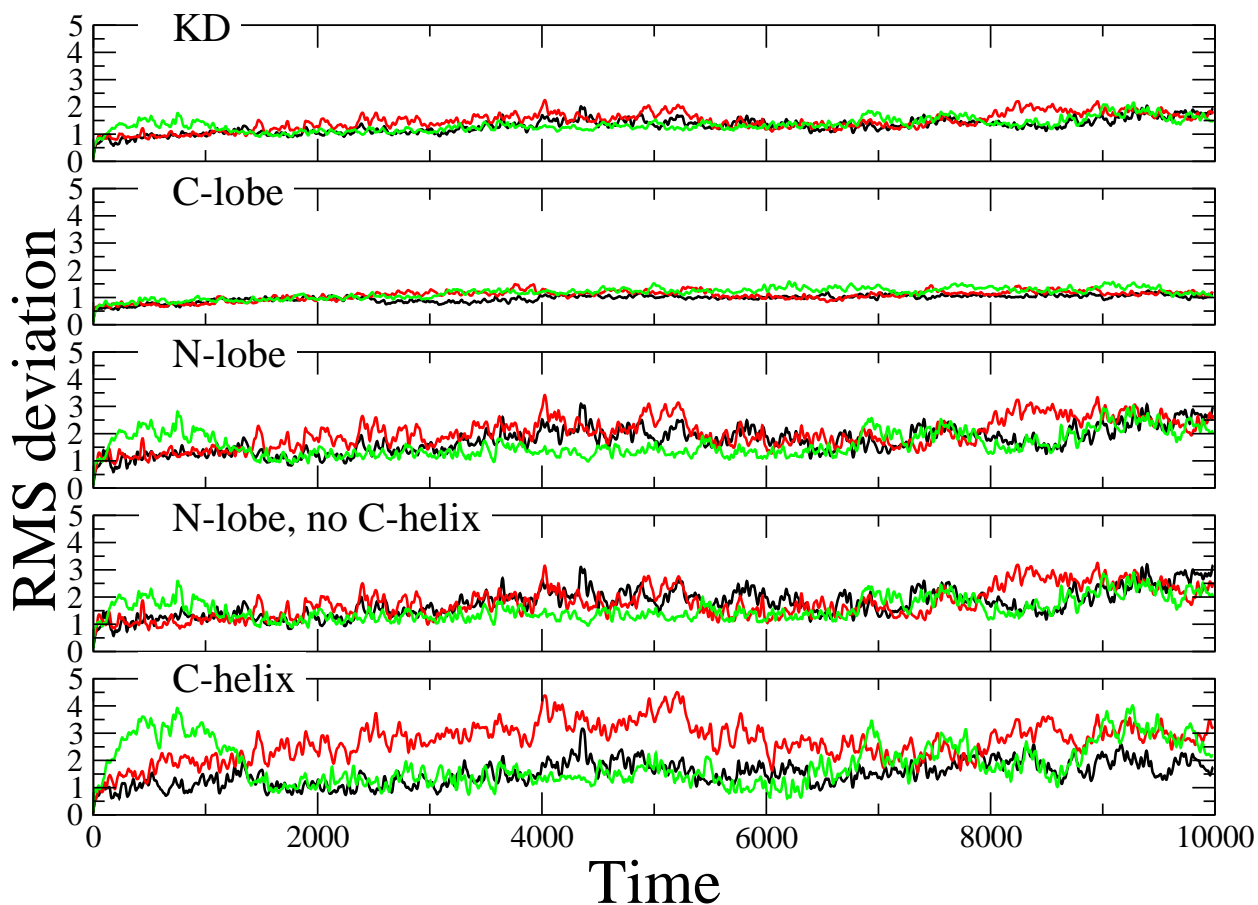
Figure 4.13: *Stability of the different Flt3 constructs. RMS deviation reported in Å are plotted as function of time in ps for the Cα-atoms of KD. Lines are drawn and colored in black for Flt3-wt, red for Flt3-593, green for Flt3-603. The RMS deviation of different protein portions are plotted in different labeled graphs.*
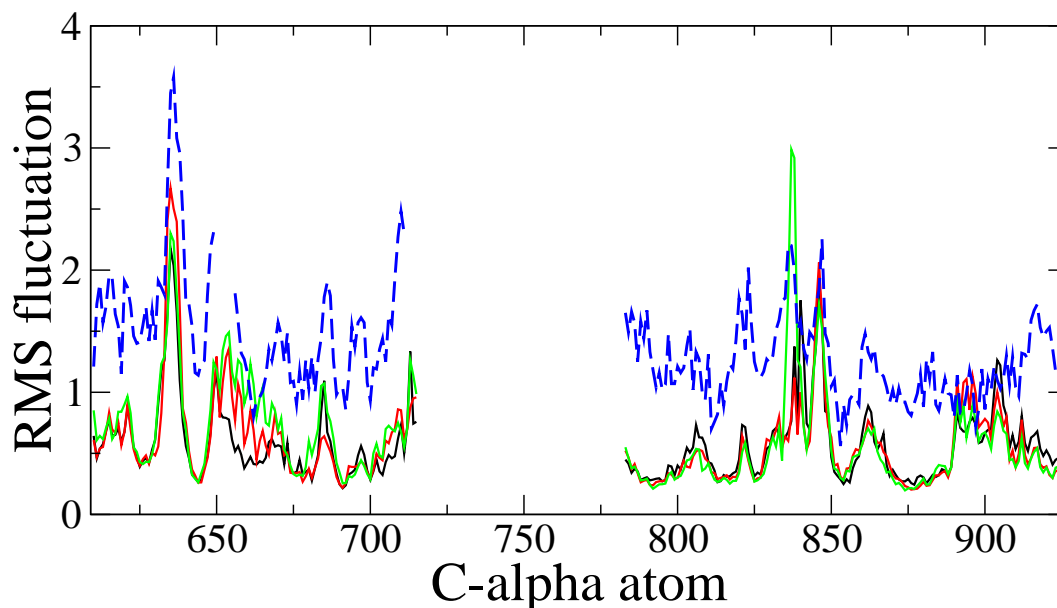
Figure 4.14: *Stability of the different Flt3 constructs. RMS fluctuation reported in Å are plotted for the residue's Cα-atom (starting from residue 609). Lines are drawn and colored in black for Flt3-wt, red for Flt3-593, green for Flt3-603. In blue dashed-line the pattern obtained from the temperature factors of the crystallographic structure, 1RJB, by applying the formula: $< R^2 >^{1/2} = (3B/8\pi^2)^{1/2}$ [261]. In the plot the missing data, revealed by the gaps in the lines, corresponds to residues of the kinase insert region which are deleted in the crystallographic structure and replaced by 4 glycines in the simulated proteins. Furthermore in the same graph some protein portions identified as residues 655-669 (C-helix), 617-622 (gly-rich loop), 631-637 (loop between β4-β5), 682-687 (loop β6-β7), 809-816 (C-loop), 829-858 (A-loop).*

values of the KD are relatively low ranging from 0.8 to 2.7 Angstrom and rather stable over time.

The "relatively low" is based on the fact that these are quite long simulations and the error summation of the algorithms used to integrate the equation of motions in MD is likely to affect the results with large amount of calculation cycles [226]. Furthermore, the lack of the pressure coupling, to generate NVT ensemble, may also affect the overall trajectory.

No major differences have been appreciated for the C-lobe with RMSD values below 1.2 Angstrom for all three proteins. In contrast, increased movement of the N-lobe have been observed for the three systems with the wild type being the more stable one. A detailed analysis clearly shows that the C-helix is indeed the structural element with the highest RMSD values and significant differences between the three systems. For Flt3-wt the deviation values range from 0.6 to 3 Angstrom, while for the two constructs with the shorter JM the values are significantly higher and fluctuate between 0.6 and 4.5 Angstrom. About the trend, Flt3-603 shows two peaks for the release from the whole juxtamembrane domain in the first nanosecond of simulation and a second deviation in the last one. Flt3-593 has a slower increase of the RMSD but a longer persistence compare to the other construct. The larger amplitude of RMS deviation of the isolated C-helix compare to the N-lobe without this element let us speculate that this helix might driven the movements of the whole N-terminal region.

The RMS fluctuations for KD Cα-atoms of the simulated wild-type protein and the ones estimate from the B-factors of the crystallographic structure were also compared (in figure 4.14 on the facing page). A qualitative agreement of the fluctuation pattern is shown for all atoms with a minor extent for the last 50 amino acids. This difference is most probably due to a less well-defined experimental electron density in this region that automatically leads to an increase in B-values. This has been often observed at the N- and C-terminus of proteins.

The comparison of the three simulated systems emphasizes the change in mobility for the atoms of the C-helix (residues 655-669) for the truncated forms respect to the wild type, with a higher amplitude for the protein without JM. The other variations in fluctuation entity are at loops like the Gly-rich loop (residues 617-622) and between the 4th and the 5th strands (631-637), and 6th and 7th ones (682-687). Furthermore, Flt3-603 compared to the other constructs has a higher mobility for some residues at the beginning of the A-loop. The N-terminal part of this loop appears to experience the conformational change of the N-lobe as depicted in figure 4.12 on page 96.

For addressing the issue about correlations between the motion of different parts of the N-lobe, the covariance matrix of the Cartesian coordinates of the Cα-atoms was calculated from the superposition of the C-lobe (described in materials and methods). To reveal the effect of the removal of the auto-inhibitory domain from the constructs, the snapshots corresponding to 1 ns period at different time points within the simulations (from 0 ns to 1 ns, from 4 ns to 5 ns and 9 ns to 10 ns) were employed separately for these analysis. The most interesting comparison between the three system evolutions was observed for the first computed nanosecond and is reported in figure 4.15 on the following page. No major motion couplings, within the analyzed time frame, were depicted for the wild type conformation. While the comparison of covariance matrices of Flt3-wt with the other constructs, shows an increment of positive couplings among N-lobe portions for

(a) *Flt3-WT (0-1 ns)*

(b) *Flt3-593 (0-1 ns)*

(c) *Flt3-603 (0-1 ns)*
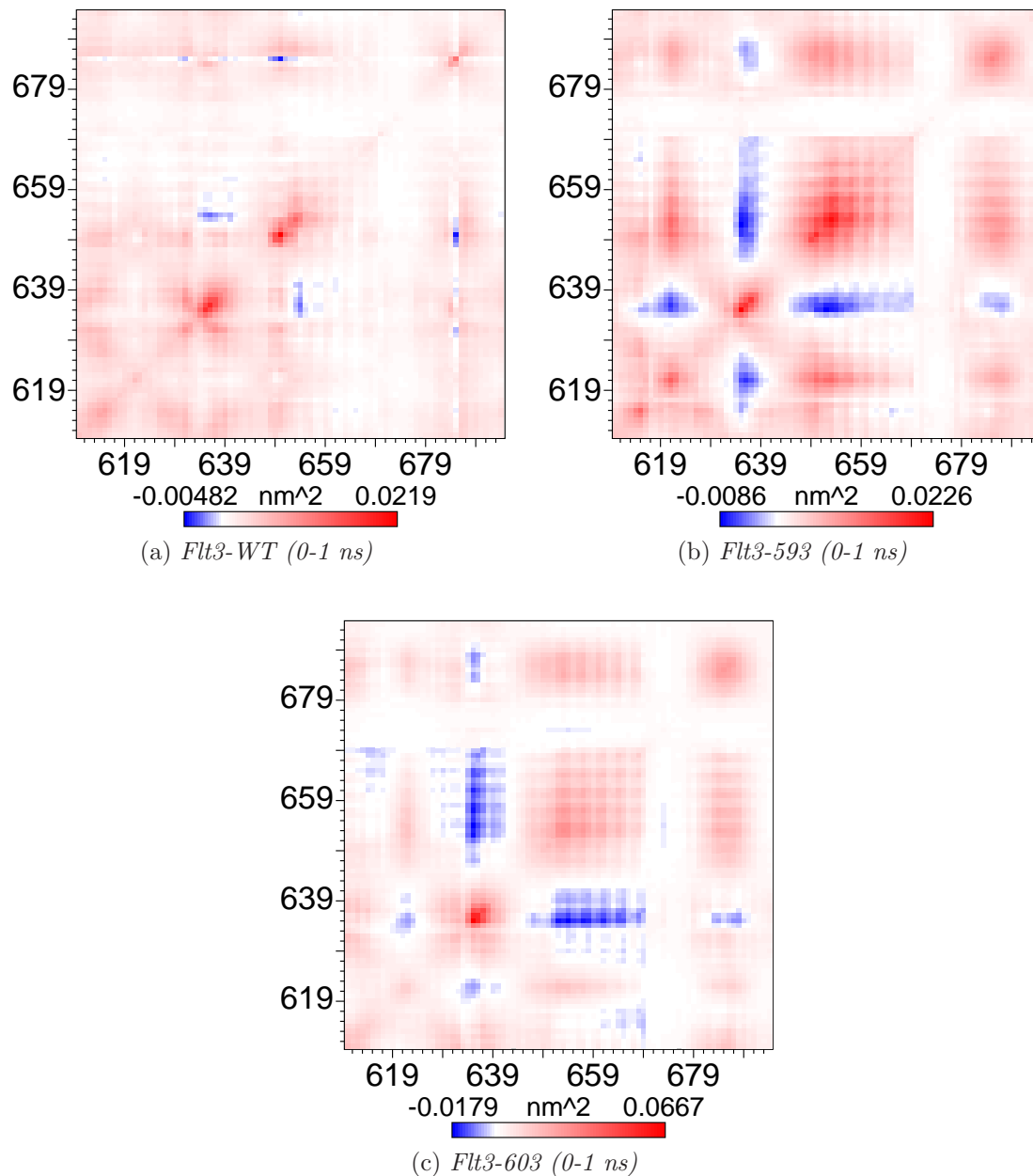
Figure 4.15: *Covariance matrices for the three simulated system. The graph demonstrates the correlation between Cα-atoms motions of N-lobe residues. The residue number is reported. Positive correlations are in red while negative ones are in blue. the color scale is associated with the normalized value of the coupling movements. Secondary structure elements of the N-lobe are identify by the residues numbers: 610-617 (β3), 622-630 (β4), 638-647 (β5), 655-669 (C-helix), 677-681 (β6) and 688-692 (β7).*

the truncated forms as well as the appearance of additional negative ones. To note, is the uncoupling maintained among movements of the loop between the C-helix and $\beta6$ (residues 669-676) and the rest of the N-lobe, as described in [247]. This element acts as a hinge for conformational changes of the small lobe due to the extended van der Waals contacts with the C-lobe. The positive coupling is appreciated for the elements on the front side of the protein (where the substrate binds) and in particular for the region encompassing residues 648 to 669 (loop C-terminal of the $\beta5$ and C-helix), 615 to 625 (part of the $\beta3$ and $\beta4$ with the loop in between) and 676 to 690 (the 6th and 7th strand and the loop in between). On the other hand, these regions have negative correlation with the movement of the loop in between the $\beta4$ and $\beta5$ which is situated on the back side of the protein. Thus, the hypothesis previously expressed about C-helix concerting the change at the N-lobe is further supported by the finding that there are correlations, both positive and negative, in the movements of the former with the ones of neighborhood protein portions.

Structural comparisons

To better understand the effect of the described movements, the visual comparison of structures at different stages of the calculated protein lifetimes and the crystallographic conformation has been performed. As anticipated by the RMSD profiles of the wild type simulation either with the first protocol (see figure 4.9 on page 93) or with the second (see figure 4.13 on page 97), changes correspond to minimal fluctuations around the experimentally determined conformation (figure 4.16a on the next page). As depicted in figure 4.12b on page 96 and figure 4.16, a fast conformational change occurs with the relief of the protein from the whole juxtamembrane, identified in the Flt3-603 constructs. The major protein substructure taking part of this deviation is recognized as the C-helix, see figure 4.13 on page 97 and in figure 4.16b. The lack of certain residues, mainly Tyr572 and Met578, due to the removal of the N-terminal portion of the autoinhibitory domain, is likely to leave this secondary structure to shift downward. The same movement is supposed to take place with the activation of the Insr and the comparison between the achieved conformation by Flt3-603 and the active and non-productive Insr is depicted in figure 4.17 on page 103. In the case where part of the JM is still present and interacts with the outer face of the C-helix (the Flt3-593 protein model) a similar rearrangement is seen but later in the evolution time (in figure 4.12a on page 96 and figure 4.16 on the following page). This delay seems to be justified by the reorganization of the hydrophobic interactions between the remaining part of the juxtamembrane and the C-helix, depicted in figure 4.18 on page 103. The Phe594 and Tyr597 partially counter balance the missing residues of the JM by their conformational rearrangements causing a different orientation of residues Tyr599 and Trp603.

The last feature extracted from the Flt3-603 trajectory is the evolving conformation of the C-helix after the conformational change. In the case of this construct, two peaks are seen for the element by the RMSD plot in figure 4.13 on page 97 suggesting a second. From the comparison of the three structure snapshots taken after the first peak (see figure 4.19 on page 104), it can be seen a sort of liberty in the movement of the helix. This element, with the N- and C-terminal loops, from the position at the 1st nanosecond is spatially translated in one direction at the 5th nanosecond and in the opposite direction at the end of the simulation, revealing the freedom of its movement.

(a) *wild-type comparison*

(b) *1st nanosecond*

(c) *5th nanosecond*

(d) *10th nanosecond*
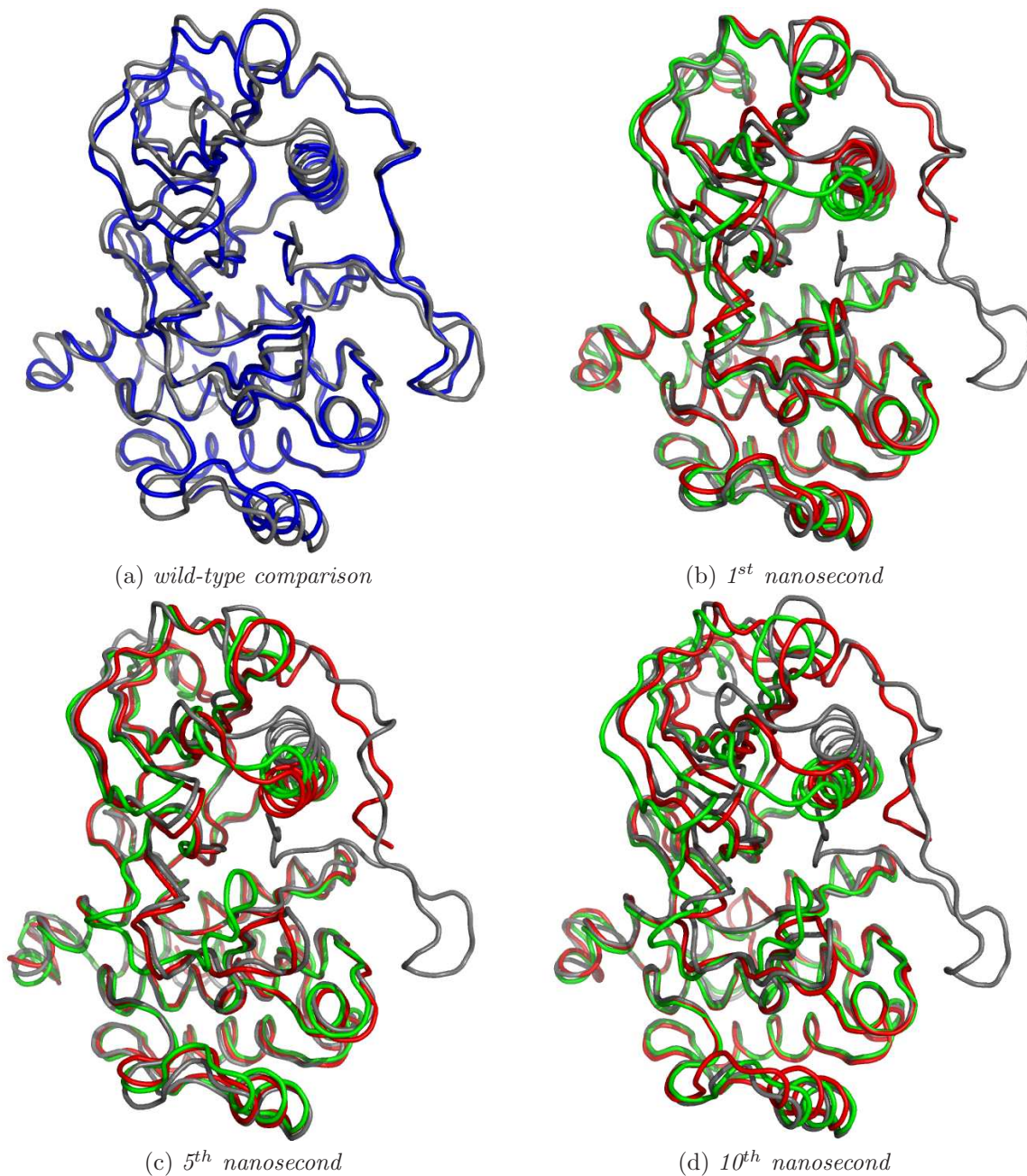
Figure 4.16: *Structural comparison of Flt3 constructs at different points of simulation. Proteins are shown with ribbon represent and color-coded. Flt3-wt is in gray, Flt3-593 is in red, Flt3-603 is in green. In (a) the comparison of Flt3-wt at 10 ns with the crystallographic structure (1RJB, in blue) is reported. A shift of the C-helices downwards can be appreciated corresponding to the movement seen for the non-productive and productive forms of Insr.*
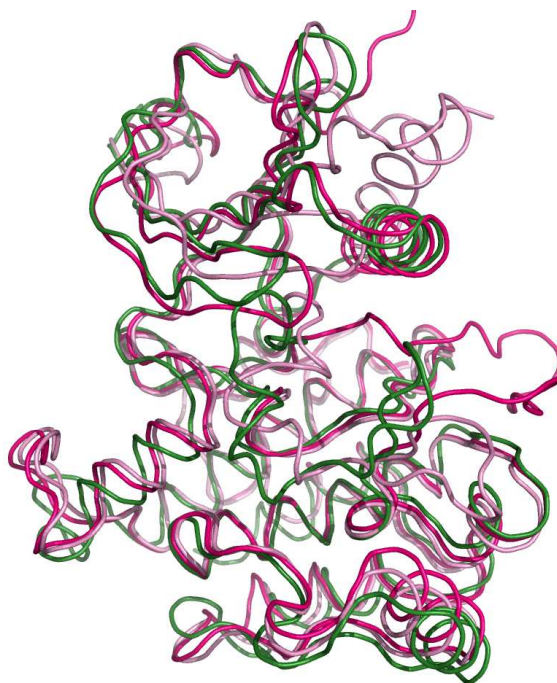
Figure 4.17: *The comparison of Flt3-603 and Insr conformation. The conformations of the Flt-603 at 10 ns, non-productive and productive states of Insr are drawn in ribbon representation and color-coded. Flt3 is in green while Insr productive form is in red and the non-productive one in pink.*
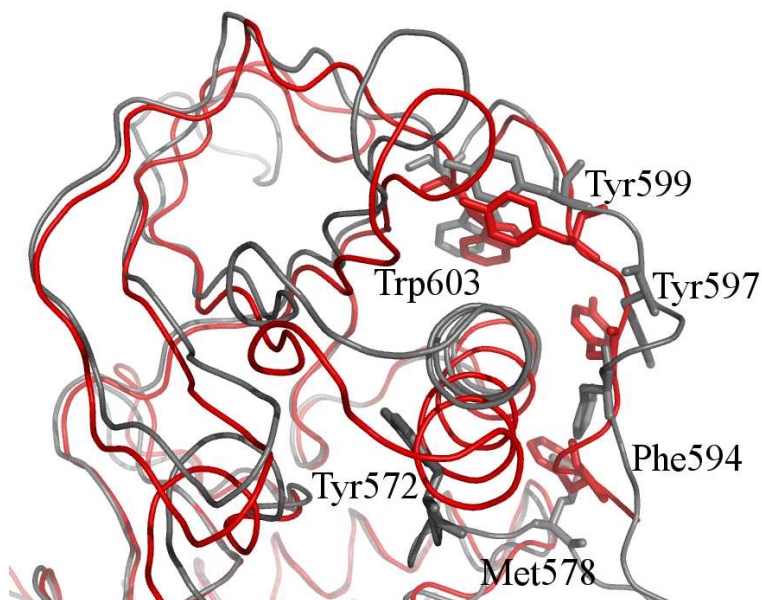
Figure 4.18: *The comparison of some interactions of JM-KD for Flt3-wt and Flt3-593. The structures taken at 10 ns of simulation are drawn in ribbon representation with some residues in sticks and labeled. Flt3-wt is in gray while Flt3-593 is in red.*
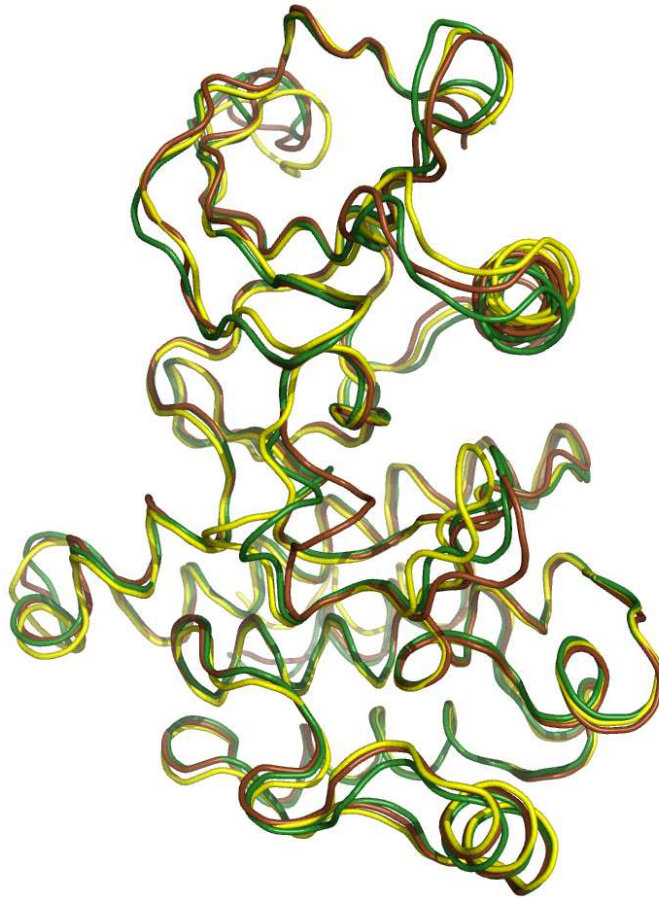
Figure 4.19: *Structural comparison of Flt3-603 at different points of simulation. The protein is shown with ribbon represent and color-coded for time: brown at 1 ns, yellow at 5 ns and green at 10 ns.*

4.5 Conclusions

The molecular mechanism/s of activation of the enzymatic moiety in tyrosine kinase proteins under either physiological or pathological conditions, is still an open issue. On the one hand, events like ligand binding, residue phosphorylation and protein-protein interactions account for the "normal" activation of these proteins. On the other hand, mutations, deletions and insertions of one or more amino acid have been correlated with their aberrant behavior, as well as protein fusion and overexpression. It is the case of Flt3 receptor tyrosine kinase where the abnormal addition of residues at the auto-inhibitory domain, the ITDs, leads to its constitutive enzymatic activation. With this study an attempt to shed light on the quest is reported.

How these insertions can affect the folding and/or the dynamic of Flt3 kinase domain? Since the auto-inhibited complex of this protein is structurally known, guesses can be made as well as trials predictions with theoretical methods. Here, different models of Flt3 bearing ITDs were generated and their dynamical behavior assessed by molecular dynamics protocols. The main information achieved by these applications is the steric hindrance experienced by the C-helix due to the juxtamembrane domain interaction.

In addition, different constructs of the JM-KD complex were tested always by means of MD to investigate the conformational changes encountered with the relief of JM portions. The C-helix appears to drive the movements of the N-lobe and to possess a certain liberty in the absence of JM interaction. The dynamics of the Flt3-593 construct is likely to reflect the beginning of the normal pathway of Flt3 activation with a slower but concerted movement of the whole N-lobe. While the behavior of the second construct might resemble more some oncogenic forms of the protein where the JM contacts are totally disrupted.

Surely, much deeper studies have to be carried out to further elucidate, prove or disprove these findings. Investigations about the larger changes that the KD undergoes for activation, namely the DFG motif switch and opening of the activation loop. Nonetheless, the results here reported let us speculate about the major role played by the C-helix during the first step for the kinase domain activation of Flt3 protein.

Chapter 5

Dynamics of residues at H1-H2 interface of Insr

5.1 Abstract

The structurally conserved regions (SCRs) in evolutionary related proteins, is an issue while applying homology modeling protocols. Their identification from the template and transposition into the target can be sometime ambiguous.

An important role for structure and dynamics of tyrosine kinase domain was attributed to a group of amino acids at the H1-H2 interface in chapter 3. Here, the investigation of this suggested SCR is done by dynamics simulations of three conformations of the Insr KD. The interface residues appeared to be stable over time with specific differences between the protein states.

Furthermore, the dynamics of the whole protein is analyzed and depicted.

5.2 Introduction

Tyrosine kinases are crucial for normal physiology of mammals [17] and are largely correlated to human malignancies [39]. Their catalytic domains share high sequence similarity and the same secondary structure disposition [77]. In addition, the activity of these enzymes, can be modulated by xenobiotics, such as Imatinib, used as cancer treatments [262]. These molecules inhibit the protein activity by competing with ATP for the same binding site [235]. Thus the kinase domain of PTKs is highly investigated as target for drug design [263].

Rational and theoretical ways to discover new medicaments came together with the knowledge and understanding of bio-macromolecules such as proteins and DNA at the molecular level [264]. In particular one of them, the so-called structure-based drug design, relies on the well refined three-dimensional structure of the putative cavity recognized as binding site [221]. The spatial arrangement of the hydrophobic and polar characters of the site, guides the researcher for finding complementary molecules whose features are likely to maximize the interaction pattern. Nowadays, the required data, namely the atomic disposition of macromolecules, can be generated experimentally or theoretically [265]. Indeed, the first way is always preferred but in some cases, when the experimental information is missing, the theoretical approach might fulfilled the need. In the case in

which the similarity between template and target is higher than 60%, the comparative model can be used as a low resolution structure (2.5-3 Angstrom) for performing structure-based design [266]. While the generation of the structure is most of the time easy, the reliability of this model remains an issue to be addressed.

"Proteins are among the most unpredictable molecules in nature" [265]. Luckily, the three-dimensional structure of proteins is more conserved than their primary sequences [267] and a small change in a protein sequence usually reflects a minor variation of its structure [268]. Carrying out comparative modeling, structurally conserved regions needs to be identified being the key elements to generate a reasonable structure [221].

Although, the first SCRs to isolate are the majority of the secondary structures, others might also be deduced by the knowledge of the protein functions and characteristics [221]. Thus, residues known to be important for catalysis, in case of enzymes, and for binding secondary partners, should correctly be assigned in the sequence alignment. In addition, certain regions of the structure can have primary roles for the general disposition and dynamics of the studied folding. The latter ones, although are sometimes difficult to infer, can be investigated by observation of structure/s in one or more states and might be confirmed by experiments such as mutational analysis. A theoretical approach, is the use of MD to study the dynamics of the examined system, which may lead to the isolation of important residues in the motion of the macromolecule.

Here, the focus is on the Insr crystallographic structures as general template for TK homology modeling. MD simulations were carried out on experimentally derived structures to assess the dynamics of the system. In addition, from chapter 3 a new feature of this protein system has been revealed; a pool of amino acids at the interface of the two lobes which appeared to be important for the relative disposition of the substructures. A particular attention is paid on the motions of these residues.

5.3   Materials and methods

The structure of the Insr kinase domain obtained by X-ray crystallography were found at the protein databank (PDB) under the entries: 1IRK, 1IR3 and 1I44. Coordinates for protein atoms and water molecules were taken from the PDB files while ligands, namely AMP-PNP and peptide substrate, and the phosphate groups of the productive conformation, 1IR3, were not considered. Furthermore, in the PDB entry 1I44 part of the A-loop (from Arg1182 to Leu1197) is missing and thus the original stretch of amino acids was added by means of MODELLER9v1. The conformation of these 16 amino acids was assigned randomly with stereochemical constrains, such as ϕ and ψ angles.

Later, water molecules were added to the systems in order to reach the total number of 12155.

The dynamics of these protein conformations were evaluated by means of GROMACS-3.3.1 software package [252–254] using settings and protocol described in chapter 3 on page 53 for NPT ensemble production. 5 nanoseconds long trajectory was simulated for each system and analyzed with the tools provided in the software package. Atomic coordinates were rendered with VMD-1.8.7 [260], BODIL-0.8.1 [61] and PYMOL [64] for visual analysis and figure generation.

5.4   Results and discussions

5.4.1   The structures

Insr in the non-productive state (PDB code: 1IRK), here alternatively called irI, possesses mainly a close conformation, with the A-loop folded in the reaction center of the enzyme, as revealed by the crystallography study [67] and biophysical measurements [245]. The DFG-motif is in the Asp-out form with the phenylalanine sandwiched by the interaction-core. A hydrogen-bond network has been described as critical for the maintenance of such conformation [92].

The kinase domain of the productive state (1IR3), also called irA, is found phosphorylated at the A-loop and in complex with a AMP-PNP molecule and a peptide substrate. The protein has DFG-motif in Asp-in form and the A-loop in a open conformation stabilized by the phosphate groups.

In addition, a third state is also present (1I44), also referred to as intermediate form or irM, where the A-loop is not completely visible but it is supposed to have an intermediate conformation between the one of irI and irA. The structure in the crystal was achieved by the replacement of Asp1138 in the A-loop to alanine [269] which resulted in the breakage of the close conformation as confirmed by other experimental results [92]. The DFG-motif is in the Asp-in conformation but the phenylalanine has a different position than in the productive form. This conformation of the protein is interesting for several reasons. Analyzing the A-loop sequences of the tyrosine kinases substitutions are frequent for residues described as part of the critical bonding for the close A-loop. Furthermore, the structure should represent an intermediate conformation between the non-productive and the productive forms of Insr and could be a valid target for drug design purpose.

5.4.2   The dynamics of Insr

The MD simulation for the three systems was carried out as described in materials and methods. To have a first feeling about the evolution of the conformations, the structural comparisons between the initial protein conformations (0 ns) and the snapshot taken at 3 nanoseconds of the trajectories, is shown in figure 5.1 on the next page. The major conformational change of all three structures is the rotation of the N-lobe on the C-lobe. The structures of irA and irM rotate in the direction of the C-helix while the irI on the opposite, toward the Gly-rich loop. In more details, the irA state appears to counteract the removal of the AMP-PNP ligand with close-up movements of the ATP-binding site. For the A-loop, both non-productive and productive KD show a stable conformation of this structural element over time. While, the intermediate state shows an appreciable extension of the A-loop which optimizes the packing with the C-lobe during simulation.

RMSD patterns

From the RMSD comparison of the Cα-atoms of the three KD states, plotted in figure 5.2 on page 111, the intermediate conformation shows to reach the dynamic equilibrium with a higher initial (within the first nanosecond) deviation compared to the other structures. Then the amplitude of fluctuations decreases with a mean value of approximately 2.5 Å for the last two nanoseconds. This value is just slightly higher than the average for the irI

(a) *non-productive state, asp-out*

(b) *productive state, asp-in*

(c) *intermediate state, asp-out*
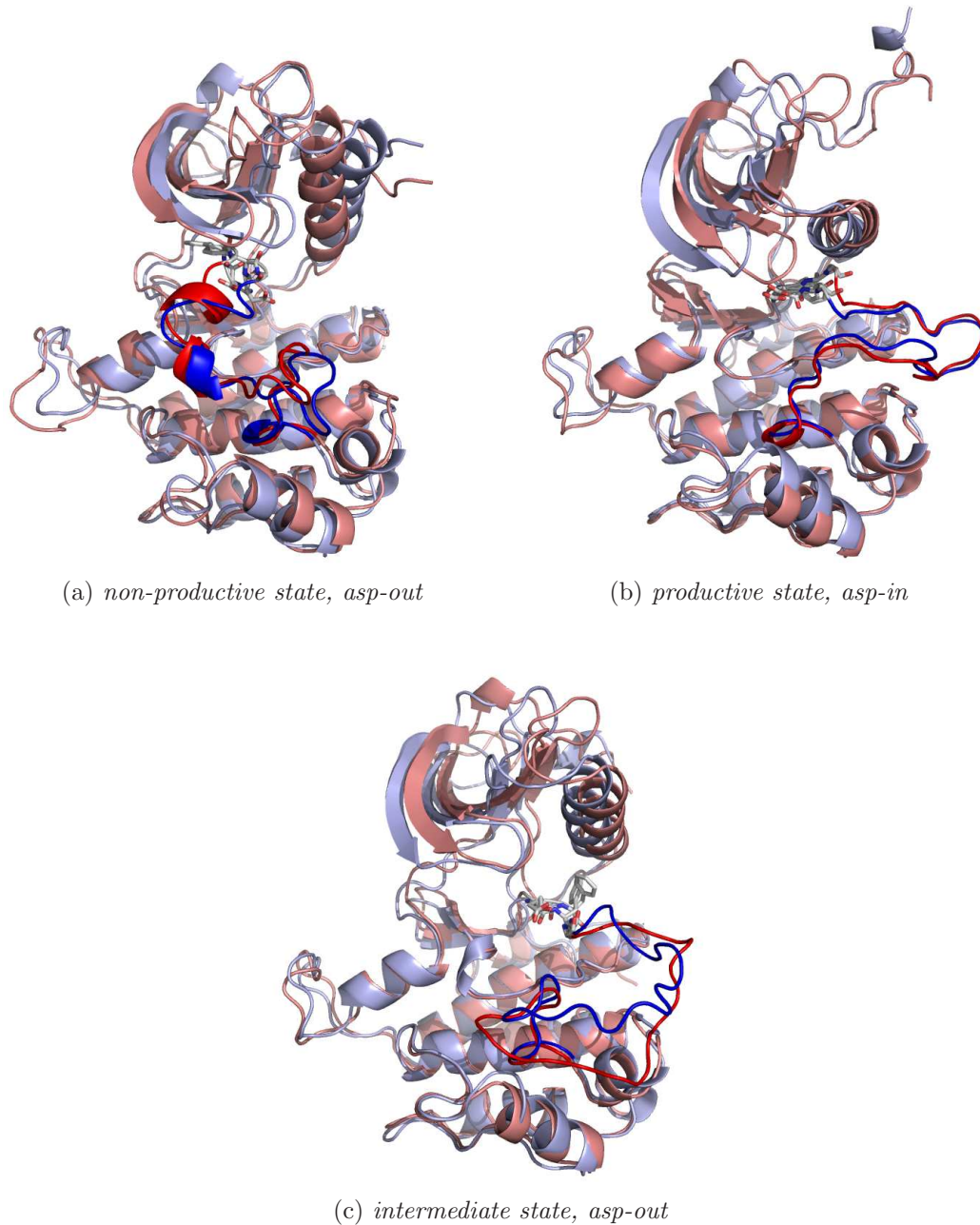
Figure 5.1: *Structural comparison at different simulation time points (0 and 3 ns) for the three KD conformations of Insr. Protein are shown with cartoon representation and color-coded. Furthermore, DFG motifs are depicted in stick representation. Structures at simulation time 0 ns (before MD) are colored in light-blue with A-loop in blue while those at time point 3 ns are colored in salmon with A-loop in red.*

(2.2 Å) and for irA (2.1 Å). When the RMS deviation is decomposed for the two lobes, as in the case of Flt3 in chapter 4 on page 79, the N-terminal substructure reaches stability with RMSD values of 1 Å higher that the one of the C-lobe. Further analysis of the N-lobe elements reveals that in contrary to the oncogenic Flt3 trends, major variations are seen for the N-lobe without C-helix than for the isolated helix. Larger fluctuations are found for the last three plots in figure 5.2, with two peaks reaching 6 Å of deviation for irM, at 0.5 and 2.5 nanoseconds, and a rather stable progression for the other two.
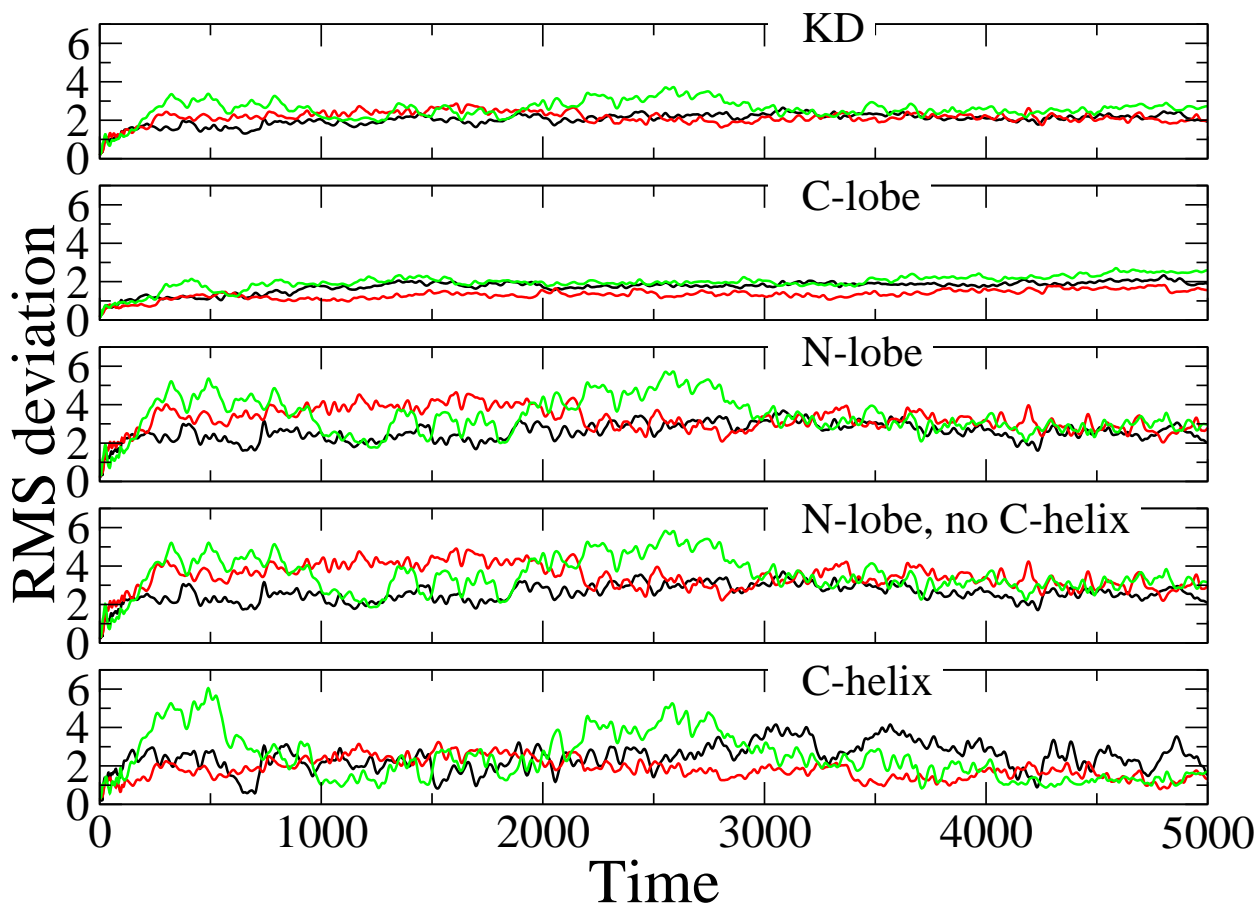
Figure 5.2: *Plots of RMS deviations of Cα-atoms of the catalytic domain for the three Insr conformations. The progressions of the RMS deviation from the initial structures is reported in Å as function of time (in ps). The non-productive, productive and intermediate states are in black, red and green, respectively.*

Other parameters

While computing the sterical parameters for the three states of the protein, an initial peak is revealed in both non-productive and intermediate conformations when compared to the stable trend acquired for the third state (see figure 5.3a and figure 5.3b on the next page). Thus, the average values are taken for the last 3 nanoseconds.

About the compactness of the structures, the progress of the radius of gyration during the simulation (figure 5.3a on the following page) shows a more packed conformation

for the irA (18,9 Å) than irI (19.2 Å) and irM (19.7 Å). The profiles of the previous patterns are in agreement with the ones observed for the inter-lobe distances. The last 3 nanoseconds show quite stable patterns with an average value of 24.2 Å for irA, 26 Å for irI and 26.8 Å for irM.

In general, all three structures are stable over time with constant folding and adjustments of the directionality of the two lobes. The portions of the N-lobe move together as a unique subdomain.

The non-productive conformation is stable but the pocket available for binding is reduced being the space occupied by the phenylalanine of the DFG-motif. As could be expected, the relaxation of productive state leads to a slight collapse of the ATP-binding site. The third structure of the Insr, the intermediate, is also stable within the simulation time. The conformation of the DFG-motif seems to help the kinase to maintained a wide-open structure with less contacts between the inner faces of the lobes.

(a) *Insr radius of gyration*   (b) *Insr inter-lobe distances*
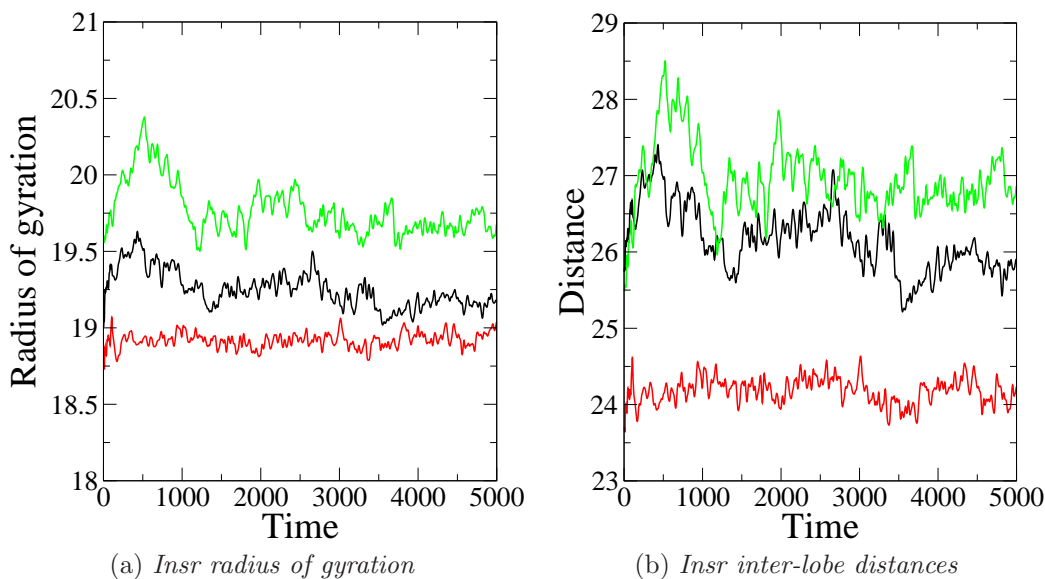
Figure 5.3: *Development of the radius of gyration and inter-lobe distance for KD conformations of Insr during simulations. The progressions of the radius of gyration (a) and of the distance between the centers of mass of the N-lobe and C-lobe (b), are reported in Å and plotted as function of time, in ps. The non-productive, productive and intermediate states are in black, red and green, respectively.*

5.4.3   Dynamics of H1-H2 interfaces

The trajectories were further analyzed to infer the dynamics of the residues at the interface of N- and C-lobe identified in chapter 3 on page 53. This pool of amino acids were suggested to play an important role for the main conformation of the kinase domain as well as for its dynamics.

As early described, changes in the van der Waals contacts are seen comparing the non-productive and productive states. A minor change for the Met1106 is appreciated from the figure 5.4 on the facing page, as well as the rearrangements of Phe1081, Leu1150 and
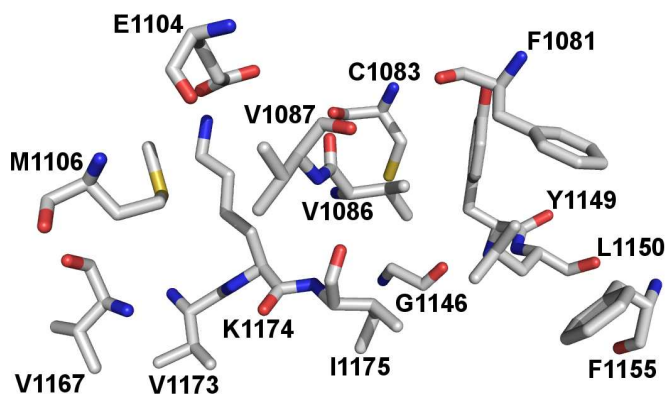
(a) *non-productive state*
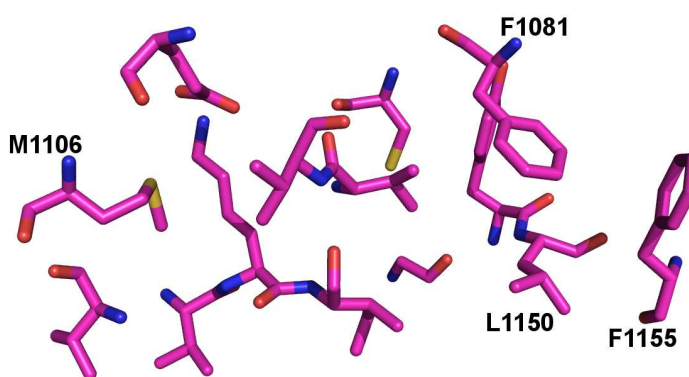
(b) *productive state*
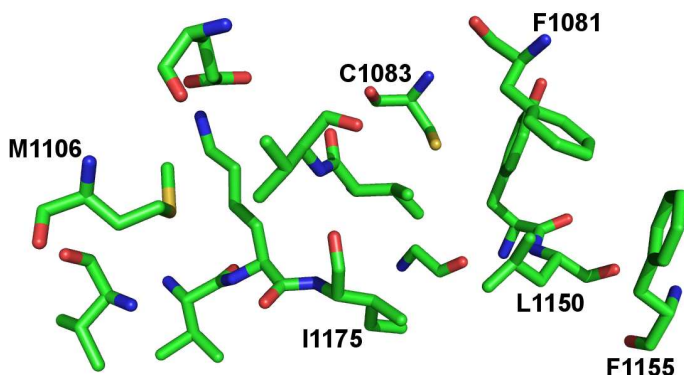
(c) *intermediate state*

Figure 5.4: *H1-H2 interface residues of the three Insr states. Amino acids at the interface of H1 and H2, according to chapter 3 on page 53, are depicted in sticks representation. The atoms are colored according to the type: blue for nitrogen, red for oxygen and yellow for sulfur. Carbon atoms are colored gray for the non-productive state, magenta for the productive and green for the intermediate. Residues are fully labelled for the non-productive state. Residues experiencing major side chain movements compared to the non-productive state are labelled in (b) and (c).*

Phe1155. Furthermore, a different disposition of certain residues is appreciated for the intermediate form (figure 5.4 on the previous page). In this state Met1106 and Leu1150 have conformations close to the ones found for the non-productive structure, while phenylalanines 1081 and 1155 are closer to the conformations of the productive one. With respect to the other two structures, Cys1083 and Ile1175 of irM show a new unrevealed movement. The side chain of the cysteine residue switches in between the one of Tyr1149 and Val1086, pushing them apart.

The dynamical behavior of the residues at the interface was firstly visually inspected with the trajectory representations in VMD. Two snapshots at different time points were extract from the simulations and superimposed on residues: Val1173, K1174 and Ile1175. This superimposition is shown in figure 5.5 on the facing page. To better appreciate the movements of the side chains of these residues at the interfaces during the three simulations, the RMS fluctuation (RMSF) of their atoms is plotted in figure 5.6 on page 116. With the relaxation of the structure in the non-productive state (irI), the Cys1083 accommodates its side chain sulfur atom between the Tyr1149 and Val1086 pushing apart these two residues. The achieved conformation (shown in figure 5.5a on the facing page) is similar to the one seen in the intermediate state before the simulation (see figure 5.4c on the previous page) and appears to trigger the second main structural modification involving Phe1081 and Phe1155. These two phenylalanines rearrange the packing of their rings with major changes expressed in the RMS fluctuation pattern (see figure 5.6 on page 116). The rest of the residues are found quite stable for irI.

In the productive state, the conformation of the residues is stable to a lesser extent compare to the irI previously described. Although their hydrophobic interactions are conserved more than in the non-productive state, the packing becomes less tight from the beginning of the simulation and is revealed by the translation of the residues around Val1173, K1174 and Ile1175 in figure 5.5b on the next page. This behavior might be connected with the removal of the ligand especially for the interaction between the ligand-core and the interaction-core. The observation would also agree with the findings that insulin receptor is more stable in the non-productive state than in the productive one, when tested in the absence of the ligand [245]. In this conformation, three side chains show larger mobility than the others, Leu1150, Val1173 and the Cys1083 which for a short period (less than 1 nanosecond), later in the simulation, switches the side chain as for the non-productive form.

In the last system analyzed, the intermediate state, the side chain movements of the two phenylalanines, 1081 and 1155, is comparable with the non-productive state, where these two plus Leu1150, continuously change interaction pattern. The Cys1083 switching, described above, is frequently seen within the trajectory. In this system, larger fluctuations are depicted for the Met1106 atoms. The motions of this residue appeared to be coupled with Val1087 through the van der Waals contacts involving their side chains.

In general, the residues at the interfaces of H1 and H2 showed fluctuations of their side chain atoms lower than 2 Å except for Phe1155 (see figure 5.6 on page 116). The polar interactions involving main chain atoms, oxygen of Ile1175 and nitrogen of Val1087, and side chain atoms, amino group of Lys1174 and carboxyl group of Glu1104, are conserved for the three evolving conformations. The apolar interaction patterns are also well conserved but present some variations like the side chain switching of Cys1083 and the continuous rearrangements of Phe1081, Leu1150 and F1155. Moreover, the intermediate

(a) *non-productive state*

(b) *productive state*

(c) *intermediate state*
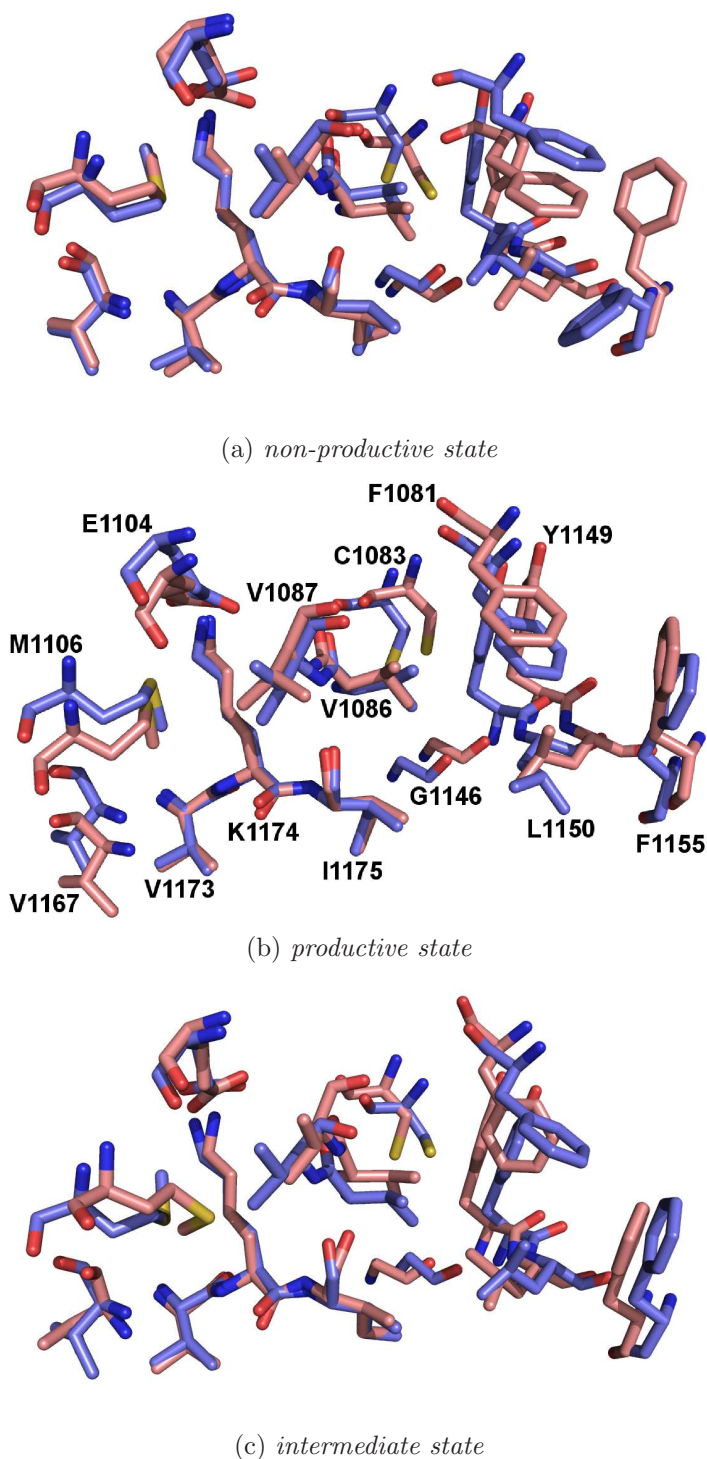
Figure 5.5: *Structural comparison of H1-H2 interfaces at different simulation time points (0 and 3 ns) for the three KD conformations of Insr. Residues are depicted in stick representation and color-coded for atoms. Nitrogen is blue, oxygen is red, sulfur is yellow and carbon is light-blue for the conformation at 0 ns and light-red for the one at 3 ns. Labels can be found for the productive state. The structures in the pictures are superposed on the central residues, Val1173, K1174 and Ile1175.*
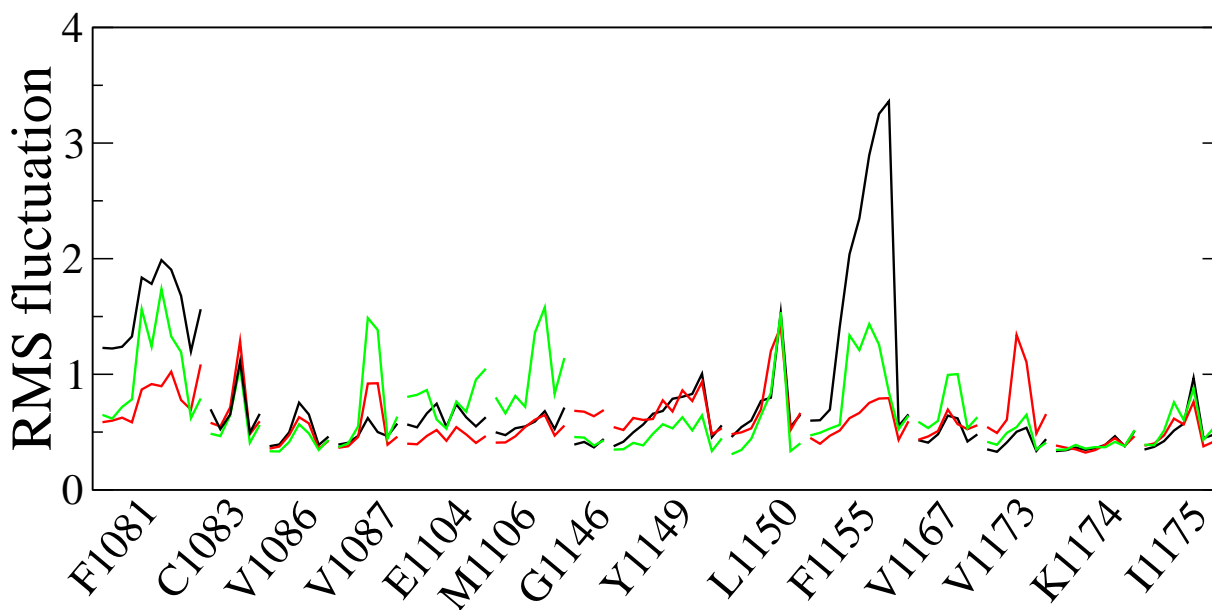
Figure 5.6: *RMS fluctuations for residues at the H1-H2 interface of Insr states. The patterns of fluctuations, in Å, are shown for atoms of the residues and color-coded for non-productive, black, productive, red, and intermediate, green. RMSF values at the extremities of each residue profile are the ones of the main chain atoms while those in between are from the side chain atoms.*

state revealed a new reorganization of Met1106 and Val1087 side chains.

Thus the packed conformation of the interface residues was found quite stable over the short period of simulation. The small variations of their hydrophobic contacts may account for larger movements in the rest of the kinase domain on the time scale of the natural dynamics. The switching of the Cys1083 could be peculiar for the Insr subfamily whereas, at the same position, an histidine residue is highly conserved. The hydrophobic contacts of Met1106 might also play a role in different dynamical behavior compared to other TKs, because this residue is often replaced by cysteines or alanines whose shorter side chains may account for these differences.

Recently, solved crystal structures of Fgfr2 KD were published for the wild-type protein as well as for some mutants [270]. The authors suggested the existence of a "molecular brake" for the kinase domain: an hydrogen-bond network that involves residues Asn549 (His1085 in Insr), Glu565 (Glu1104 in Insr) and Lys641 (Lys1174 in Insr). Glutamate and lysine are among those residues identified at the H1-H2 interface group. *In-vitro* assays of the Fgfr2 mutants showed their higher activities compare to the wild-type protein which is in agreement with the finding that they occur in certain malignancies [270]. Moreover, the structural comparison reveal the modification of molecular brake conformation for mutants affecting these residues as well as for the productive state of the protein.

The reported results of the mutations of Fgfr2, at the position of Glu1104 and Lys1174 for Insr, support the critical role played by one of the two polar interactions of the interface suggested in chapter 3 on page 53.

Thus, the next arisen issue, is the major function of the other residues at the interface

of the two lobes, which are forming the tight packing suggested to rule the structure disposition and dynamics of TK catalytic domains.
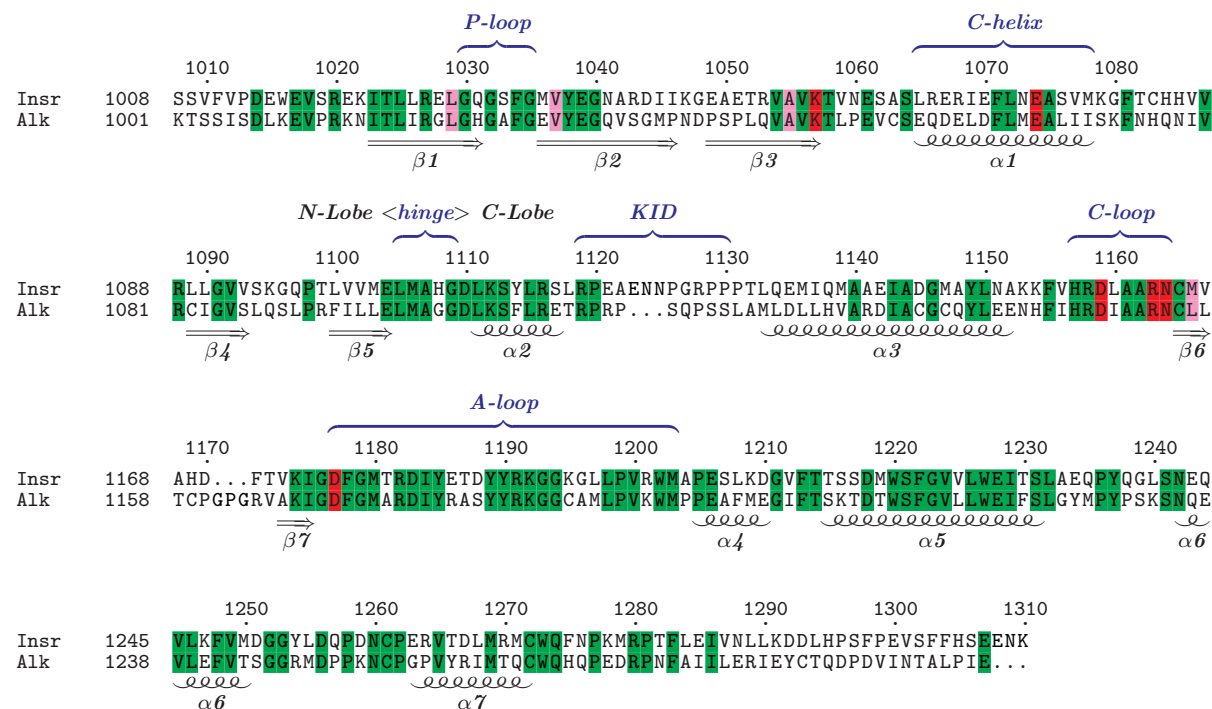
Figure 5.7: *Sequence alignment of Insr and Alk KDs. The amino acid sequences of the two TKs are shown in one-letter code, secondary structures are labeled at the bottom while defined regions are depicted at the top and colored in blue. All conserved residues of the sequences are colored in green, residues important for catalysis are in red and residues binding to the ligand core are shown in light magenta. Insr KD sequence corresponding to the crystallographic structure 1IRK is shown while the Alk sequence as well as the numbering for both come from* EXPASY *entries (P06213 for Insr and Q9UM73 for Alk.)*

A Hint about this subject came from the results obtained in a previous study about Alk kinase domain (data not shown). Homology models were carried out based on the three structures of Insr and the protein conformations were optimized via MD simulation. The models and their dynamics were judged sub-optimal because the side chain of methionine at the hinge region, corresponding to Met1106 in Insr, was modelled in a different conformation than found in the template structures. The difference is the result of the construction of the loop between $\beta6$ and $\beta7$ which is longer in the target sequence than in the template (in the previous alignment). This element is found close in space to the methionine, in the predicted structure, and its conformation would clash with the methionine side chain as found in the template. The reason was attributed to the presence of two proline within the sequence of the loop (see alignment above). Thus, it is likely that this presence restrains the spatial conformation of the loop because of the restricted ϕ and ψ angle values achievable by this type of residue. The packing of the residues at H1-H2 interface of the protein models resulted different than seen for Insr. Although the overall folding was maintained, the dynamical evolution of the system deviates from the initial coordinates with large changes for the methionine and the protein.

This was interpreted as a confirmation of the importance of the correct position of methionine residue.

5.5 Conclusions

Having in mind homology modeling for structure-based drug design, the Insr crystallographic structure 1IR3 is an ideal candidate to produce models of the kinase domain in the active state. The relaxed structure is less suitable because the out coming conformation from MD depicts a smaller ATP-binding site due to the missing interaction of binding-core. The intermediate conformation is surely an interesting structure for protein modeling and its ATP-binding pocket has a particular shape given by the conformation of the DFG-motif and the A-loop. The binding-core residues do not interact with any partner and are free for new contacts. Initial coordinates and their snapshots at different time points of the calculated trajectory might be used to enlarge the set of templates for intermediate protein conformations.

In respect to the H1-H2 interface subject, the investigations here reported confirmed the hypothesis of the central role of this region in the KD features of Insr. The answer to certain questions like: which are the structural and dynamical effects for the replacement of hydrophobic residues of this region? and again, are there large variations of their interaction patterns while KD undergoes activation?, might be of great help for a better understanding of protein kinases under normal and aberrant conditions. These issues could be addressed by combining site-directed mutagenesis, fluorescence, spectroscopy and MD studies on mutated KDs. In addition, further studies might extend these findings by investigation of structure and dynamics of others tyrosine kinases which are predicted to highlight difference for the kinase domain features.

Final conclusions

Studies concerning sequence, structure and dynamics of tyrosine kinase domains were presented in the chapters of this thesis. One of the problems occurring when the subject of the study is a large protein family, as in the case of tyrosine kinases, is the wealth of information to deal with. Although, there is no doubt regarding the usefulness of the thousands of publications on the topic, the collection of the complete set of information can be an issue.

While carrying out these studies more than hundred solved crystal structures, with a large portion released just in the last 2 years, were collected and investigated. Furthermore, a protocol for the alignment of the 90 KD sequences of TKs is given but the out-coming result[1] might be difficult to exploit because of the laboriousness to display it on normal computer screens or A4 paper format with a readable font size. The used solution was to print parts of the alignment in different sheets and then join them together, in a "leporello-like" way.

With the comparison of the KD structures in complex with ligands, the minimal requirements for a molecule to bind at the ATP pocket were identified and called "ligand core". One hydrogen bond with the hinge region and several hydrophobic contacts with 4 residues, whose side chains have a conserved hydrophobic character throughout the kinases, was recognized as the "interaction core". These knowledge can be used as minimal pharmacophore model to screen molecules against the cofactor pocket for hit finding during drug design processes.

With the aim of finding reasons at the sequence level for conformational differences, further comparison of the kinase domain crystals have been performed. In this respect, 5 clusters of mainly hydrophobic residues were identified and their disposition in different states and kinases investigated. These residues interact in different ways depending on the KD type and state. Thus, families of kinases were suggested because characterized by similar arrangements of the clusters in non-productive and productive conformations.

A deeper observation of two of the clusters, namely the hydrophobic core of the two lobes, revealed a special configuration of 14 residues at the interface. On the one hand, in the inner part of the protein, some of these residues build up two facing planes whose function is suggested to allow the independent movements of the two lobes. On the other hand, the outer part of this interface depicts two pawls from the C-lobe which embed the N-lobe by surrounding residues of the loop at C-terminal position of the C-helix. The stability of these residues was explored by MD simulations of experimental structures of the non-productive, productive and intermediate states of Insr. Major fluctuations are seen for residues on one side of the interface, closer to the C-helix, in all the conformations

---

[1]Attached as appendix in soft form within the compact disk

119

of the KD. Furthermore, other amino acids at the interface showed different dynamical behavior characterizing the states of the protein. Replacements of the less conserved H1-H2 interface residues are predicted to characterize the KD conformation and dynamics of members of the TK family. In a possible panorama, deduced by these findings, the interactions among the interface residues change while the protein undergoes from one state to another one. Such changes are partly ruled by the C-helix element through its interaction with some interface residues, at the position of Phe1081 and Phe1155 in Insr, and its influence on the conformation of the loop found at its C-terminal position.

This piece of information can also be useful within homology modeling for correct assignment of this structurally conserved region in the sequence alignment and its proper conformation in the predicted structure. The latter could be further optimized by structure relaxation with MD simulations.

The last issue analyzed is the role of C-helix in the activation of Flt3 kinase domain. This protein was depicted by X-ray crystallography in complex with its juxtamembrane domain which inhibits the kinase activity stabilizing the non-productive state of the KD. Aberrant structural modifications of the auto-inhibitory domain have been correlated with the constitutive activation of Flt3 and found in the occurrence of certain hematopoiesis. Normal and oncogenic ways of activation of this kinase were predicted to follow the same path where the interaction between JM and the enzyme is disrupted.

The comparison of the dynamics of Flt3 constructs, with shorter JMs, and wild-type protein revealed the function of the auto-inhibitory domain blocking the movements of the C-helix. With the relief of the juxtamembrane, the N-lobe undergoes structural changes mainly driven by the C-helix. Thus, the role of the element in the first step of the KD activation of both wild type and oncogenic proteins was suggested.

Earlier findings about the ligand-complexes focused on defining 38 residues which built up the ATP-binding site. The amino acid variance among protein kinases was pointed out and major efforts were spent on how to achieve the ligand-binding selectivity. Here, the function of the binding-core residues is revealed crucial for stabilizing the KD conformation. The binding-core of tyrosine kinases can interact with the phenylalanine of the DFG-motif in the KD apo-form, while it is always in contact with the ligand-core in the bound-form. Thus, the ligand-core is seen as the minimal pharmacophore for TK inhibitors competing with ATP.

Several studies on the KD structures showed how interaction and conformation of three elements of the KD, namely C-helix, Gly-rich loop and A-loop, rule the structural arrangement of specific tyrosine kinases in certain states. In the present work a different point of view was used to understand a large set of conformations of different TKs. The hypothesis is that there should be stretches of amino acids in the protein primary structure of KDs whose variations account for the different three-dimensional topologies as depicted by the experimental structures. The study revealed the H1-H2 interface residues as possible region implicated in the stabilization of the achievable conformations and in the dynamics of the catalytic domain. Although, it is not the only region to rule such features, it is for sure an important and until now less explored one.

The molecular description of the activation of kinase domain have been already attempt before. In these studies, usually, conformational changes are forced by the application of external forces or potentials to the system. The novelty of the trials here presented, is the simulation of a perturbed auto-inhibitory system. The response of the Flt3 KD to

the release of its auto-inhibitory mechanism was monitored by MD. The interest was to highlight which elements of the protein experience motion while the system goes from one thermodynamic state toward another one. The approach was successfully and resulted in the C-helix dynamics playing a major role for the N-lobe conformational change. Which is suggested as first step in the activation partway.

Bibliography

[1] W Hendriks, J Leunissen, E Nevo, H Bloemendal, and W W de Jong. The lens protein alpha a-crystallin of the blind mole rat, spalax ehrenbergi: evolutionary change and functional constraints. *Proc Natl Acad Sci U S A*, 84(15):5320–5324, Aug 1987.

[2] N Iwabe, K Kuma, and T Miyata. Evolution of gene families and relationship with organismal evolution: rapid divergence of tissue-specific genes in the early evolution of chordates. *Mol Biol Evol*, 13(3):483–493, Mar 1996.

[3] W E Müller. Origin of metazoan adhesion molecules and adhesion receptors as deduced from cdna analyses in the marine sponge geodia cydonium: a review. *Cell Tissue Res*, 289(3):383–395, Sep 1997.

[4] S L Baldauf. The deep roots of eukaryotes. *Science*, 300(5626):1703–1706, Jun 2003.

[5] H Suga, K Kuma, N Iwabe, N Nikoh, K Ono, M Koyanagi, D Hoshiyama, and T Miyata. Intermittent divergence of the protein tyrosine kinase family during animal evolution. *FEBS Lett*, 412(3):540–546, Aug 1997.

[6] H Suga, M Koyanagi, D Hoshiyama, K Ono, N Iwabe, K Kuma, and T Miyata. Extensive gene duplication in the early evolution of animals before the parazoan-eumetazoan split demonstrated by g proteins and protein tyrosine kinases from sponge and hydra. *J Mol Evol*, 48(6):646–653, Jun 1999.

[7] A Goffeau, B G Barrell, H Bussey, R W Davis, B Dujon, H Feldmann, F Galibert, J D Hoheisel, C Jacq, M Johnston, E J Louis, H W Mewes, Y Murakami, P Philippsen, H Tettelin, and S G Oliver. Life with 6000 genes. *Science*, 274(5287):563–567, Oct 1996.

[8] The Washington University Genome Sequencing Center. Genome sequence of the nematode c. elegans: a platform for investigating biology. *Science*, 282(5396):2012–2018, Dec 1998.

[9] D W Meinke, J M Cherry, C Dean, S D Rounsley, and M Koornneef. Arabidopsis thaliana: a model plant for genome analysis. *Science*, 282(5389):679–682, Oct 1998.

[10] S Y Rhee, S Weng, D Flanders, J M Cherry, C Dean, C Lister, M Anderson, M Koornneef, D W Meinke, T Nickle, K Smith, and S D Rounsley. Genome maps 9. arabidopsis thaliana. wall chart. *Science*, 282(5389):663–667, Oct 1998.

[11] M D Adams, S E Celniker, R A Holt, C A Evans, J D Gocayne, P G Amanatides, S E Scherer, P W Li, R A Hoskins, and et al. The genome sequence of drosophila melanogaster. *Science*, 287(5461):2185–2195, Mar 2000.

[12] J C Venter, M D Adams, E W Myers, P W Li, R J Mural, G G Sutton, H O Smith, M Yandell, C A Evans, R A Holt, J D Gocayne, and et al. The sequence of the human genome. *Science*, 291(5507):1304–1351, Feb 2001.

[13] T Miyata and H Suga. Divergence pattern of animal gene families and relationship with the cambrian explosion. *Bioessays*, 23(11):1018–1027, Nov 2001.

[14] S H Shiu and W H Li. Origins, lineage-specific expansions, and multiple losses of tyrosine kinases in eukaryotes. *Mol Biol Evol*, 21(5):828–840, May 2004.

[15] J Gu and X Gu. Natural history and functional divergence of protein tyrosine kinases. *Gene*, 317(1-2):49–57, Oct 2003.

[16] H Suga, K Katoh, and T Miyata. Sponge homologs of vertebrate protein tyrosine kinases and frequent domain shufflings in the early evolution of animals before the parazoan-eumetazoan split. *Gene*, 280(1-2):195–201, Dec 2001.

[17] T Hunter. The croonian lecture 1997. the phosphorylation of proteins on tyrosine: its role in cell growth and disease. *Philos Trans R Soc Lond B Biol Sci*, 353(1368):583–605, Apr 1998.

[18] J L Tan and J A Spudich. Developmentally regulated protein-tyrosine kinase genes in dictyostelium discoideum. *Mol Cell Biol*, 10(7):3578–3583, Jul 1990.

[19] N King and S B Carroll. A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. *Proc Natl Acad Sci U S A*, 98(26):15032–15037, Dec 2001.

[20] G Schieven, J Thorner, and G S Martin. Protein-tyrosine kinase activity in saccharomyces cerevisiae. *Science*, 231(4736):390–393, Jan 1986.

[21] R A Lindberg, A M Quinn, and T Hunter. Dual-specificity protein kinases: will any hydroxyl do? *Trends Biochem Sci*, 17(3):114–119, Mar 1992.

[22] P van der Geer, T Hunter, and R A Lindberg. Receptor protein-tyrosine kinases and their signal transduction pathways. *Annu Rev Cell Biol*, 10:251–337, 1994.

[23] H W Mewes, K Albermann, M Bähr, D Frishman, A Gleissner, J Hani, K Heumann, K Kleine, A Maierl, S G Oliver, F Pfeiffer, and A Zollner. Overview of the yeast genome. *Nature*, 387(6632 Suppl):7–65, May 1997.

[24] S A Chervitz, L Aravind, G Sherlock, C A Ball, E V Koonin, S S Dwight, M A Harris, K Dolinski, S Mohr, T Smith, S Weng, J M Cherry, and D Botstein. Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science*, 282(5396):2022–2028, Dec 1998.

[25] E Barizza, F Lo Schiavo, M Terzi, and F Filippini. Evidence suggesting protein tyrosine phosphorylation in plants depends on the developmental conditions. *FEBS Lett*, 447(2-3):191–194, Mar 1999.

[26] K Kameyama, Y Kishi, M Yoshimura, N Kanzawa, M Sameshima, and T Tsuchiya. Tyrosine phosphorylation in plant bending. *Nature*, 407(6800):37–37, Sep 2000.

[27] D Miranda-Saavedra and G J Barton. Classification and functional annotation of eukaryotic protein kinases. *Proteins*, Jun 2007.

[28] Y Kishi, C Clements, D C Mahadeo, D A Cotter, and M Sameshima. High levels of actin tyrosine phosphorylation: correlation with the dormant state of dictyostelium spores. *J Cell Sci*, 111 (Pt 19):2923–2932, Oct 1998.

[29] N King, C T Hittinger, and S B Carroll. Evolution of key cell signaling and adhesion protein families predates animal origins. *Science*, 301(5631):361–363, Jul 2003.

[30] Y Segawa, H Suga, N Iwabe, C Oneyama, T Akagi, T Miyata, and M Okada. Functional development of src tyrosine kinases during evolution from a unicellular ancestor to multicellular animals. *Proc Natl Acad Sci U S A*, 103(32):12021–12026, Aug 2006.

[31] B C Leadbeater. Life-history and ultrastructure of a new marine species of proterospongia (choanoflagellida). *Journal of Marine Biology Association*, 63(1):135–160, 1983.

[32] D J Hibberd. Observations on the ultrastructure of the choanoflagellate codosiga botrytis (ehr.) saville-kent with special reference to the flagellar apparatus. *J Cell Sci*, 17(1):191–219, Jan 1975.

[33] E Avizienyte, A W Wyke, R J Jones, G W McLean, M A Westhoff, V G Brunton, and M C Frame. Src-induced de-regulation of e-cadherin in colon cancer cells requires integrin signalling. *Nat Cell Biol*, 4(8):632–638, Aug 2002.

[34] W Rengifo-Cam, A Konishi, N Morishita, H Matsuoka, T Yamori, S Nada, and M Okada. Csk defines the ability of integrin-mediated cell adhesion and migration in human colon cancer cells: implication for a potential role in cancer metastasis. *Oncogene*, 23(1):289–297, Jan 2004.

[35] T M Chiang, J Reizer, and E H Beachey. Serine and tyrosine protein kinase activities in streptococcus pyogenes. phosphorylation of native and synthetic peptides of streptococcal m proteins. *J Biol Chem*, 264(5):2957–2962, Feb 1989.

[36] T Hunter. Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell*, 80(2):225–236, Jan 1995.

[37] T Hunter. A thousand and one protein kinases. *Cell*, 50(6):823–829, Sep 1987.

[38] G Manning, D B Whyte, R Martinez, T Hunter, and S Sudarsanam. The protein kinase complement of the human genome. *Science*, 298(5600):1912–1934, Dec 2002.

[39] P Blume-Jensen and T Hunter. Oncogenic kinase signalling. *Nature*, 411(6835):355–365, May 2001.

[40] G Burnett and E P Kennedy. The enzymatic phosphorylation of proteins. *J Biol Chem*, 211(2):969–980, Dec 1954.

[41] E G Krebs, D J Graves, and E H Fischer. Factors affecting the activity of muscle phosphorylase b kinase. *J Biol Chem*, 234:2867–2873, Nov 1959.

[42] D A Walsh, J P Perkins, and E G Krebs. An adenosine 3',5'-monophosphate-dependant protein kinase from rabbit skeletal muscle. *J Biol Chem*, 243(13):3763–3765, Jul 1968.

[43] E G Krebs and J A Beavo. Phosphorylation-dephosphorylation of enzymes. *Annu Rev Biochem*, 48:923–959, 1979.

[44] S Shoji, K Titani, J G Demaille, and E H Fischer. Sequence of two phosphorylated sites in the catalytic subunit of bovine cardiac muscle adenosine 3':5'-monophosphate-dependent protein kinase. *J Biol Chem*, 254(14):6211–6214, Jul 1979.

[45] B E Kemp, R B Pearson, V Guerriero, I C Bagchi, and A R Means. The calmodulin binding domain of chicken smooth muscle myosin light chain kinase contains a pseudosubstrate sequence. *J Biol Chem*, 262(6):2542–2548, Feb 1987.

[46] L W Slice and S S Taylor. Expression of the catalytic subunit of camp-dependent protein kinase in escherichia coli. *J Biol Chem*, 264(35):20940–20946, Dec 1989.

[47] W Eckhart, M A Hutchinson, and T Hunter. An activity phosphorylating tyrosine in polyoma t antigen immunoprecipitates. *Cell*, 18(4):925–933, Dec 1979.

[48] T Hunter and B M Sefton. Transforming gene product of rous sarcoma virus phosphorylates tyrosine. *Proc Natl Acad Sci U S A*, 77(3):1311–1315, Mar 1980.

[49] O N Witte, A Dasgupta, and D Baltimore. Abelson murine leukaemia virus protein is phosphorylated in vitro to form phosphotyrosine. *Nature*, 283(5750):826–831, Feb 1980.

[50] T Hunter and J A Cooper. Protein-tyrosine kinases. *Annu Rev Biochem*, 54:897–930, 1985.

[51] S K Hanks, A M Quinn, and T Hunter. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science*, 241(4861):42–52, Jul 1988.

[52] S K Hanks and T Hunter. Protein kinases 6. the eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J*, 9(8):576–596, May 1995.

[53] D R Knighton, J H Zheng, L F Ten Eyck, V A Ashford, N H Xuong, S S Taylor, and J M Sowadski. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*, 253(5018):407–414, Jul 1991.

[54] D R Knighton, J H Zheng, L F Ten Eyck, N H Xuong, S S Taylor, and J M Sowadski. Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*, 253(5018):414–420, Jul 1991.

[55] T Pawson. Protein modules and signalling networks. *Nature*, 373(6515):573–580, Feb 1995.

[56] H Yamaguchi, M Matsushita, A C Nairn, and J Kuriyan. Crystal structure of the atypical protein kinase domain of a trp channel with phosphotransferase activity. *Mol Cell*, 7(5):1047–1057, May 2001.

[57] N LaRonde-LeBlanc and A Wlodawer. Crystal structure of a. fulgidus rio2 defines a new family of serine protein kinases. *Structure*, 12(9):1585–1594, Sep 2004.

[58] D Drennan and A G Ryazanov. Alpha-kinases: analysis of the family and comparison with conventional protein kinases. *Prog Biophys Mol Biol*, 85(1):1–32, May 2004.

[59] S K Hanks and A M Quinn. Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol*, 200:38–62, 1991.

[60] D G Higgins and P M Sharp. Clustal: a package for performing multiple sequence alignment on a microcomputer. *Gene*, 73(1):237–44, 1988.

[61] J V Lehtonen, D J Still, V V Rantanen, J Ekholm, D Björklund, Z Iftikhar, M Huhtala, S Repo, A Jussila, J Jaakkola, O Pentikäinen, T Nyrönen, T Salminen, M Gyllenberg, and M S Johnson. Bodil: a molecular modeling environment for structure-function analysis and drug design. *J Comput Aided Mol Des*, 18(6):401–419, Jun 2004.

[62] E Beitz. Texshade: shading and labeling of multiple sequence alignments using latex2 epsilon. *Bioinformatics*, 16(2):135–139, Feb 2000.

[63] N Kannan and A F Neuwald. Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component? *J Mol Biol*, 351(5):956–972, Sep 2005.

[64] W L DeLano. The pymol molecular graphics system. *http//www.pymol.org/*, 2003.

[65] D A Johnson, P Akamine, E Radzio-Andzelm, M Madhusudan, and S S Taylor. Dynamics of camp-dependent protein kinase. *Chem Rev*, 101(8):2243–2270, Aug 2001.

[66] R A Engh and D Bossemeyer. Structural aspects of protein kinase control-role of conformational flexibility. *Pharmacol Ther*, 93(2-3):99–111, Feb-Mar 2002.

[67] S R Hubbard, L Wei, L Ellis, and W A Hendrickson. Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature*, 372(6508):746–754, Dec 1994.

[68] S R Hubbard. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and atp analog. *EMBO J*, 16(18):5572–5581, Sep 1997.

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

[70] M A Seeliger, B Nagar, F Frank, X Cao, M N Henderson, and J 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, Mar 2007.

[71] P D Jeffrey, A A Russo, K Polyak, E Gibbs, J Hurwitz, J Massagué, and N P Pavletich. Mechanism of cdk activation revealed by the structure of a cyclina-cdk2 complex. *Nature*, 376(6538):313–320, Jul 1995.

[72] X Zhang, J Gureasko, K Shen, P A Cole, and J Kuriyan. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell*, 125(6):1137–1149, Jun 2006.

[73] C D Mol, D R Dougan, T R Schneider, R J Skene, M L Kraus, D N Scheibe, G P Snell, H Zou, B C Sang, and K P Wilson. Structural basis for the autoinhibition and sti-571 inhibition of c-kit tyrosine kinase. *J Biol Chem*, 279(30):31655–31663, Jul 2004.

[74] A Vulpetti and R Bosotti. Sequence and structural analysis of kinase atp pocket residues. *Farmaco*, 59(10):759–765, Oct 2004.

[75] L Moretti, L Tchernin, and L Scapozza. Tyrosine kinase drug discovery: what can be learned from solved crystal structures? *ARKIVOC*, 2006(8):1443–1445, May 2006.

[76] L N Johnson, E D Lowe, M E Noble, and D J Owen. The eleventh datta lecture. the structural basis for substrate recognition and control by protein kinases. *FEBS Lett*, 430(1-2):1–11, Jun 1998.

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

[78] S S Taylor, J Yang, J Wu, N M Haste, E Radzio-Andzelm, and G Anand. Pka: a portrait of protein kinase dynamics. *Biochim Biophys Acta*, 1697(1-2):259–269, Mar 2004.

[79] I Tsigelny, J P Greenberg, S Cox, W L Nichols, S S Taylor, and L F Ten Eyck. 600 ps molecular dynamics reveals stable substructures and flexible hinge points in camp dependent protein kinase. *Biopolymers*, 50(5):513–524, Oct 1999.

[80] A Cavalli, C Dezi, G Folkers, L Scapozza, and M Recanatini. Three-dimensional model of the cyclin-dependent kinase 1 (cdk1): Ab initio active site parameters for molecular dynamics studies of cdks. *Proteins*, 45(4):478–485, Dec 2001.

[81] 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):115–126, Apr 2001.

[82] B Nagar, O Hantschel, M A Young, K Scheffzek, D Veach, W Bornmann, B Clarkson, G Superti-Furga, and J Kuriyan. Structural basis for the autoinhibition of c-abl tyrosine kinase. *Cell*, 112(6):859–871, Mar 2003.

[83] B Lu, C F Wong, and J A McCammon. Release of adp from the catalytic subunit of protein kinase a: a molecular dynamics simulation study. *Protein Sci*, 14(1):159–168, Jan 2005.

[84] C P Barrett and M E Noble. Molecular motions of human cyclin-dependent kinase 2. *J Biol Chem*, 280(14):13993–14005, Apr 2005.

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

[86] Y Cheng, Y Zhang, and J A McCammon. How does activation loop phosphorylation modulate catalytic activity in the camp-dependent protein kinase: a theoretical study. *Protein Sci*, 15(4):672–683, Apr 2006.

[87] B D Grant, I Tsigelny, J A Adams, and S S Taylor. Examination of an active-site electrostatic node in the camp-dependent protein kinase catalytic subunit. *Protein Sci*, 5(7):1316–1324, Jul 1996.

[88] C T Kong and P F Cook. Isotope partitioning in the adenosine 3',5'-monophosphate dependent protein kinase reaction indicates a steady-state random kinetic mechanism. *Biochemistry*, 27(13):4795–4799, Jun 1988.

[89] J Shaffer and J A Adams. An atp-linked structural change in protein kinase a precedes phosphoryl transfer under physiological magnesium concentrations. *Biochemistry*, 38(17):5572–5581, Apr 1999.

[90] J A Adams and S S Taylor. Energetic limits of phosphotransfer in the catalytic subunit of camp-dependent protein kinase as measured by viscosity experiments. *Biochemistry*, 31(36):8516–8522, Sep 1992.

[91] J Lew, N Coruh, I Tsigelny, S Garrod, and S S Taylor. Synergistic binding of nucleotides and inhibitors to camp-dependent protein kinase examined by acrylodan fluorescence spectroscopy. *J Biol Chem*, 272(3):1507–1513, Jan 1997.

[92] A J Ablooglu and R A Kohanski. Activation of the insulin receptor's kinase domain changes the rate-determining step of substrate phosphorylation. *Biochemistry*, 40(2):504–513, Jan 2001.

[93] J Shaffer and J A Adams. Detection of conformational changes along the kinetic pathway of protein kinase a using a catalytic trapping technique. *Biochemistry*, 38(37):12072–12079, Sep 1999.

[94] F Viñals, M Camps, X Testar, M Palacín, and A Zorzano. Effect of cations on the tyrosine kinase activity of the insulin receptor: inhibition by fluoride is magnesium dependent. *Mol Cell Biochem*, 171(1-2):69–73, Jun 1997.

[95] S A Lieser, B E Aubol, L Wong, P A Jennings, and J A Adams. Coupling phosphoryl transfer and substrate interactions in protein kinases. *Biochim Biophys Acta*, 1754(1-2):191–199, Dec 2005.

[96] P A Cole, D Sondhi, and K Kim. Chemical approaches to the study of protein tyrosine kinases and their implications for mechanism and inhibitor design. *Pharmacol Ther*, 82(2-3):219–229, May-Jun 1999.

[97] K Parang and P A Cole. Designing bisubstrate analog inhibitors for protein kinases. *Pharmacol Ther*, 93(2-3):145–157, Feb-Mar 2002.

[98] J Zhou and J A Adams. Is there a catalytic base in the active site of camp-dependent protein kinase? *Biochemistry*, 36(10):2977–2984, 1997.

[99] J C Hart, I H Hillier, N A Burton, and D W Sheppard. An alternative role for the conserved asp residue in phosphoryl transfer reactions. *J Am Chem Soc*, 120(51):13535–13536, Dec 1998.

[100] J C Hart, D W Sheppard, I H Hillier, and N A Burton. What is the mechanism of phosphoryl transfer in protein kinases? a hybrid quantum mechanical/molecular mechanical study. *Chem Commun*, 1:79–80, 1999.

[101] M C Hutter and V Helms. Influence of key residues on the reaction mechanism of the camp-dependent protein kinase. *Protein Sci*, 8(12):2728–2733, Dec 1999.

[102] D W Sheppard, N A Burton, and I H Hillier. The emerging normal and disease-related roles of anaplastic lymphoma kinase. *THEOCHEM*, 506(1-3):35–44, Jul 2000.

[103] Y Hirano, M Hata, T Hoshino, and M Tsuda. Quantum chemical study on the catalytic mechanism of the c-subunit of camp-dependent protein kinase. *J Phys Chem B*, 106(22):5788–5792, May 2002.

[104] K Ginalski, P Grochowski, B Lesyng, and D Shugar. Dft calculations and parametrization of the approximate valence bond method to describe the phosphoryl transfer reaction in a model system. *Int J Quantum Chem*, 90:1129–1139, 2002.

[105] M Valiev, R Kawai, J A Adams, and J H Weare. The role of the putative catalytic base in the phosphoryl transfer reaction in a protein kinase: first-principles calculations. *J Am Chem Soc*, 125(33):9926–9927, Aug 2003.

[106] N Díaz and M J Field. Insights into the phosphoryl-transfer mechanism of camp-dependent protein kinase from quantum chemical calculations and molecular dynamics simulations. *J Am Chem Soc*, 126(2):529–542, Jan 2004.

[107] G Henkelman, M X LaBute, C S Tung, P W Fenimore, and B H McMahon. Conformational dependence of a protein kinase phosphate transfer reaction. *Proc Natl Acad Sci U S A*, 102(43):15347–15351, Oct 2005.

[108] A Cavalli, M De Vivo, and M Recanatini. Density functional study of the enzymatic reaction catalyzed by a cyclin-dependent kinase. *Chem Commun (Camb)*, (11):1308–1309, Jun 2003.

[109] Y Cheng, Y Zhang, and J A McCammon. How does the camp-dependent protein kinase catalyze the phosphorylation reaction: An ab initio qm/mm study. *J Am Chem Soc*, 127(5):1553–1562, Jan 2005.

[110] M De Vivo, A Cavalli, P Carloni, and M Recanatini. Computational study of the phosphoryl transfer catalyzed by a cyclin-dependent kinase. *Chemistry*, Jul 2007.

[111] H L De Bondt, J Rosenblatt, J Jancarik, H D Jones, D O Morgan, and S H Kim. Crystal structure of cyclin-dependent kinase 2. *Nature*, 363(6430):595–602, Jun 1993.

[112] D R Knighton, J H Zheng, L F Ten Eyck, V A Ashford, N H Xuong, S S Taylor, and J M Sowadski. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*, 253(5018):407–414, Jul 1991.

[113] S H Hu, M W Parker, J Y Lei, M C Wilce, G M Benian, and B E Kemp. Insights into autoregulation from the crystal structure of twitchin kinase. *Nature*, 369(6481):581–584, Jun 1994.

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

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

[116] I Moarefi, M LaFevre-Bernt, F Sicheri, M Huse, C H Lee, J Kuriyan, and W T Miller. Activation of the src-family tyrosine kinase hck by sh3 domain displacement. *Nature*, 385(6617):650–653, Feb 1997.

[117] J Griffith, J Black, C Faerman, L Swenson, M Wynn, F Lu, J Lippke, and K Saxena. The structural basis for autoinhibition of flt3 by the juxtamembrane domain. *Mol Cell*, 13(2):169–178, Jan 2004.

[118] B M Sefton, T Hunter, K Beemon, and W Eckhart. Evidence that the phosphorylation of tyrosine is essential for cellular transformation by rous sarcoma virus. *Cell*, 20(3):807–816, Jul 1980.

[119] M A Snyder, J M Bishop, J P McGrath, and A D Levinson. A mutation at the atp-binding site of pp60v-src abolishes kinase activity, transformation, and tumorigenicity. *Mol Cell Biol*, 5(7):1772–1779, Jul 1985.

[120] M P Kamps and B M Sefton. Neither arginine nor histidine can carry out the function of lysine-295 in the atp-binding site of p60src. *Mol Cell Biol*, 6(3):751–757, Mar 1986.

[121] H Ushiro and S Cohen. Identification of phosphotyrosine as a product of epidermal growth factor-activated protein kinase in a-431 cell membranes. *J Biol Chem*, 255(18):8363–8365, Sep 1980.

[122] M Kasuga, F A Karlsson, and C R Kahn. Insulin stimulates the phosphorylation of the 95,000-dalton subunit of its own receptor. *Science*, 215(4529):185–187, Jan 1982.

[123] B Ek, B Westermark, A Wasteson, and C H Heldin. Stimulation of tyrosine-specific phosphorylation by platelet-derived growth factor. *Nature*, 295(5848):419–420, Feb 1982.

[124] M J Eck, S Dhe-Paganon, T Trüb, R T Nolte, and S E Shoelson. Structure of the irs-1 ptb domain bound to the juxtamembrane region of the insulin receptor. *Cell*, 85(5):695–705, May 1996.

[125] C Ortutay, J Väliaho, K Stenberg, and M Vihinen. Kinmutbase: a registry of disease-causing mutations in protein kinase domains. *Hum Mutat*, 25(5):435–442, May 2005.

[126] C B Gambacorti-Passerini, R H Gunby, R Piazza, A Galietta, R Rostagno, and L Scapozza. Molecular mechanisms of resistance to imatinib in philadelphia-chromosome-positive leukaemias. *Lancet Oncol*, 4(2):75–85, Feb 2003.

[127] D R Robinson, Y M Wu, and S F Lin. The protein tyrosine kinase family of the human genome. *Oncogene*, 19(49):5548–5557, Nov 2000.

[128] A Ullrich and J Schlessinger. Signal transduction by receptors with tyrosine kinase activity. *Cell*, 61(2):203–212, Apr 1990.

[129] S R Hubbard. Autoinhibitory mechanisms in receptor tyrosine kinases. *Frontiers in Bioscience*, 7:1553–1562d330–340, Feb 2002.

[130] S R Hubbard. Structural analysis of receptor tyrosine kinases. *Prog Biophys Mol Biol*, 71(3-4):343–358, 1999.

[131] L E Wybenga-Groot, B Baskin, S H Ong, J Tong, T Pawson, and F Sicheri. Structural basis for autoinhibition of the ephb2 receptor tyrosine kinase by the unphosphorylated juxtamembrane region. *Cell*, 106(6):745–757, Sep 2001.

[132] K L Binns, P P Taylor, F Sicheri, T Pawson, and S J Holland. Phosphorylation of tyrosine residues in the kinase domain and juxtamembrane region regulates the biological and catalytic activities of eph receptors. *Mol Cell Biol*, 20(13):4791–4805, Jul 2000.

[133] S Mori, L Rönnstrand, K Yokote, A Engström, S A Courtneidge, L Claesson-Welsh, and C H Heldin. Identification of two juxtamembrane autophosphorylation sites in the pdgf beta-receptor; involvement in the interaction with src family tyrosine kinases. *EMBO J*, 12(6):2257–2264, Jun 1993.

[134] R M Baxter, J P Secrist, R R Vaillancourt, and A Kazlauskas. Full activation of the platelet-derived growth factor beta-receptor kinase involves multiple events. *J Biol Chem*, 273(27):17050–17055, Jul 1998.

[135] R Herbst and S J Burden. The juxtamembrane region of musk has a critical role in agrin-mediated signaling. *EMBO J*, 19(1):67–77, Jan 2000.

[136] Y Yarden and A Ullrich. Growth factor receptor tyrosine kinases. *Annu Rev Biochem*, 57:443–478, 1988.

[137] H E Tornqvist and J Avruch. Relationship of site-specific beta subunit tyrosine autophosphorylation to insulin activation of the insulin receptor (tyrosine) protein kinase activity. *J Biol Chem*, 263(10):4593–4601, Apr 1988.

[138] L M Shewchuk, A M Hassell, B Ellis, W D Holmes, R Davis, E L Horne, S H Kadwell, D D McKee, and J T Moore. Structure of the tie2 rtk domain: self-inhibition by the nucleotide binding loop, activation loop, and c-terminal tail. *Structure*, 8(11):1105–1113, Nov 2000.

[139] X L Niu, K G Peters, and C D Kontos. Deletion of the carboxyl terminus of tie2 enhances kinase activity, signaling, and function. evidence for an autoinhibitory mechanism. *J Biol Chem*, 277(35):31768–31773, Aug 2002.

[140] C Greenfield, I Hiles, M D Waterfield, M Federwisch, A Wollmer, T L Blundell, and N McDonald. Epidermal growth factor binding induces a conformational change in the external domain of its receptor. *EMBO J*, 8(13):4115–4123, Dec 1989.

[141] M A Lemmon, Z Bu, J E Ladbury, M Zhou, D Pinchasi, I Lax, D M Engelman, and J Schlessinger. Two egf molecules contribute additively to stabilization of the egfr dimer. *EMBO J*, 16(2):281–294, Jan 1997.

[142] A Yayon, M Klagsbrun, J D Esko, P Leder, and D M Ornitz. Cell surface, heparin-like molecules are required for binding of basic fibroblast growth factor to its high affinity receptor. *Cell*, 64(4):841–848, Feb 1991.

[143] M Mohammadi, S K Olsen, and O A Ibrahimi. Structural basis for fibroblast growth factor receptor activation. *Cytokine Growth Factor Rev*, 16(2):107–137, Apr 2005.

[144] S Schlee, P Carmillo, and A Whitty. Quantitative analysis of the activation mechanism of the multicomponent growth-factor receptor ret. *Nat Chem Biol*, 2(11):636–644, Nov 2006.

[145] A L Stiegler, S J Burden, and S R Hubbard. Crystal structure of the agrin-responsive immunoglobulin-like domains 1 and 2 of the receptor tyrosine kinase musk. *J Mol Biol*, 364(3):424–433, Dec 2006.

[146] S R Hubbard, M Mohammadi, and J Schlessinger. Autoregulatory mechanisms in protein-tyrosine kinases. *J Biol Chem*, 273(20):11987–11990, May 1998.

[147] J Schlessinger. Cell signaling by receptor tyrosine kinases. *Cell*, 103(2):211–225, Oct 2000.

[148] H Riedel, T J Dull, A M Honegger, J Schlessinger, and A Ullrich. Cytoplasmic domains determine signal specificity, cellular routing characteristics and influence ligand binding of epidermal growth factor and insulin receptors. *EMBO J*, 8(10):2943–2954, Oct 1989.

[149] R Lammers, A Gray, J Schlessinger, and A Ullrich. Differential signalling potential of insulin- and igf-1-receptor cytoplasmic domains. *EMBO J*, 8(5):1369–1375, May 1989.

[150] J Lee, T J Dull, I Lax, J Schlessinger, and A Ullrich. Her2 cytoplasmic domain generates normal mitogenic and transforming signals in a chimeric receptor. *EMBO J*, 8(1):167–173, Jan 1989.

[151] M A McTigue, J A Wickersham, C Pinko, R E Showalter, C V Parast, A Tempczyk-Russell, M R Gehring, B Mroczkowski, C C Kan, J E Villafranca, and K Appelt. Crystal structure of the kinase domain of human vascular endothelial growth factor receptor 2: a key enzyme in angiogenesis. *Structure*, 7(3):319–330, Mar 1999.

[152] J A Escobedo and L T Williams. A pdgf receptor domain essential for mitogenesis but not for many other responses to pdgf. *Nature*, 335(6185):85–87, Sep 1988.

[153] G R Taylor, M Reedijk, V Rothwell, L Rohrschneider, and T Pawson. The unique insert of cellular and viral fms protein tyrosine kinase domains is dispensable for enzymatic and transforming activities. *EMBO J*, 8(7):2029–2037, Jul 1989.

[154] L Ellis, E Clauser, D O Morgan, M Edery, R A Roth, and W J Rutter. Replacement of insulin receptor tyrosine residues 1162 and 1163 compromises insulin-stimulated kinase activity and uptake of 2-deoxyglucose. *Cell*, 45(5):721–732, Jun 1986.

[155] H Kato, T N Faria, B Stannard, C T Roberts, and D LeRoith. Essential role of tyrosine residues 1131, 1135, and 1136 of the insulin-like growth factor-i (igf-i) receptor in igf-i action. *Mol Endocrinol*, 8(1):40–50, Jan 1994.

[156] M Mohammadi, J Schlessinger, and S R Hubbard. Structure of the fgf receptor tyrosine kinase domain reveals a novel autoinhibitory mechanism. *Cell*, 86(4):577–587, Aug 1996.

[157] N Gotoh, A Tojo, M Hino, Y Yazaki, and M Shibuya. A highly conserved tyrosine residue at codon 845 within the kinase domain is not required for the transforming activity of human epidermal growth factor receptor. *Biochem Biophys Res Commun*, 186(2):768–774, Jul 1992.

[158] P P Knowles, J Murray-Rust, S Kjaer, R P Scott, S Hanrahan, M Santoro, C F Ibáñez, and N Q McDonald. Structure and chemical inhibition of the ret tyrosine kinase domain. *J Biol Chem*, 281(44):33577–33587, Nov 2006.

[159] K L Carraway and L C Cantley. A neu acquaintance for erbb3 and erbb4: a role for receptor heterodimerization in growth signaling. *Cell*, 78(1):5–8, Jul 1994.

[160] W K Hong and A Ullrich. *The role of EGFR in solid tumors and implications for theraphy.* Number 1. Oncol Biother, 2000.

[161] S V Sharma, D W Bell, J Settleman, and D A Haber. Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer*, 7(3):169–181, Mar 2007.

[162] E Lengyel, K Sawada, and R Salgia. Tyrosine kinase mutations in human cancer. *Curr Mol Med*, 7(1):77–84, Feb 2007.

[163] M Nakao, S Yokota, T Iwai, H Kaneko, S Horiike, K Kashima, Y Sonoda, T Fujimoto, and S Misawa. Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. *Leukemia*, 10(12):1911–1918, Dec 1996.

[164] C Miething, C Peschel, and J Duyster. Targeting the oncogenic tyrosine kinase npm-alk in lymphoma: the role of murine models in defining pathogenesis and treatment options. *Curr Drug Targets*, 7(10):1329–1334, Oct 2006.

[165] A Levitzki and A Gazit. Tyrosine kinase inhibition: an approach to drug development. *Science*, 267(5205):1782–1788, Mar 1995.

[166] K Neet and T Hunter. Vertebrate non-receptor protein-tyrosine kinase families. *Genes Cells*, 1(2):147–169, Feb 1996.

[167] J E Darnell, I M Kerr, and G R Stark. Jak-stat pathways and transcriptional activation in response to ifns and other extracellular signaling proteins. *Science*, 264(5164):1415–1421, Jun 1994.

[168] J N Ihle. Cytokine receptor signalling. *Nature*, 377(6550):591–594, Oct 1995.

[169] T Hunter and G D Plowman. The protein kinases of budding yeast: six score and more. *Trends Biochem Sci*, 22(1):18–22, Jan 1997.

[170] S R Hubbard and J H Till. Protein tyrosine kinase structure and function. *Annu Rev Biochem*, 69:373–398, 2000.

[171] D Sondhi and P A Cole. Domain interactions in protein tyrosine kinase csk. *Biochemistry*, 38(34):11147–11155, Aug 1999.

[172] A Ogawa, Y Takayama, H Sakai, K T Chong, S Takeuchi, A Nakagawa, S Nada, M Okada, and T Tsukihara. Structure of the carboxyl-terminal src kinase, csk. *J Biol Chem*, 277(17):14351–14354, Apr 2002.

[173] M B Lamers, A A Antson, R E Hubbard, R K Scott, and D H Williams. Structure of the protein tyrosine kinase domain of c-terminal src kinase (csk) in complex with staurosporine. *J Mol Biol*, 285(2):713–725, Jan 1999.

[174] Z Fan and J Mendelsohn. Therapeutic application of anti-growth factor receptor antibodies. *Curr Opin Oncol*, 10(1):67–73, Jan 1998.

[175] E Zwick, J Bange, and A Ullrich. Receptor tyrosine kinases as targets for anticancer drugs. *Trends Mol Med*, 8(1):17–23, Jan 2002.

[176] M M Goldenberg. Trastuzumab, a recombinant dna-derived humanized monoclonal antibody, a novel agent for the treatment of metastatic breast cancer. *Clin Ther*, 21(2):309–318, Feb 1999.

[177] K Parang, J H Till, A J Ablooglu, R A Kohanski, S R Hubbard, and P A Cole. Mechanism-based design of a protein kinase inhibitor. *Nat Struct Biol*, 8(1):37–41, Jan 2001.

[178] C García-Echeverría, P Traxler, and D B Evans. Atp site-directed competitive and irreversible inhibitors of protein kinases. *Med Res Rev*, 20(1):28–57, Jan 2000.

[179] P Cohen. The development and therapeutic potential of protein kinase inhibitors. *Curr Opin Chem Biol*, 3(4):459–465, Aug 1999.

[180] G Scapin. Structural biology in drug design: selective protein kinase inhibitors. *Drug Discov Today*, 7(11):601–611, Jun 2002.

[181] R Capdeville, E Buchdunger, J Zimmermann, and A Matter. Glivec (sti571, imatinib), a rationally developed, targeted anticancer drug. *Nat Rev Drug Discov*, 1(7):493–502, Jul 2002.

[182] T W Traut. Physiological concentrations of purines and pyrimidines. *Mol Cell Biochem*, 140(1):1–22, Nov 1994.

[183] S Wong, J McLaughlin, D Cheng, C Zhang, K M Shokat, and O N Witte. Sole bcr-abl inhibition is insufficient to eliminate all myeloproliferative disorder cell populations. *Proc Natl Acad Sci U S A*, 101(50):17456–17461, Dec 2004.

[184] J Dancey and E A Sausville. Issues and progress with protein kinase inhibitors for cancer treatment. *Nat Rev Drug Discov*, 2(4):296–313, Apr 2003.

[185] S Kobayashi, T J Boggon, T Dayaram, P A Jänne, O Kocher, M Meyerson, B E Johnson, M J Eck, D G Tenen, and B Halmos. Egfr mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med*, 352(8):786–792, Feb 2005.

[186] A Hochhaus, S Kreil, A Corbin, P La Rosée, T Lahaye, U Berger, N C Cross, W Linkesch, B J Druker, R Hehlmann, C Gambacorti-Passerini, G Corneo, and M D'Incalci. Roots of clinical resistance to sti-571 cancer therapy. *Science*, 293(5538):2163–2163, Sep 2001.

[187] W Matthews, C T Jordan, G W Wiegand, D Pardoll, and I R Lemischka. A receptor tyrosine kinase specific to hematopoietic stem and progenitor cell-enriched populations. *Cell*, 65(7):1143–1152, Jun 1991.

[188] O Rosnet, S Marchetto, O deLapeyriere, and D Birnbaum. Murine flt3, a gene encoding a novel tyrosine kinase receptor of the pdgfr/csf1r family. *Oncogene*, 6(9):1641–1650, Sep 1991.

[189] O Rosnet, C Schiff, M J Pébusque, S Marchetto, C Tonnelle, Y Toiron, F Birg, and D Birnbaum. Human flt3/flk2 gene: cdna cloning and expression in hematopoietic cells. *Blood*, 82(4):1110–1119, Aug 1993.

[190] D Small, M Levenstein, E Kim, C Carow, S Amin, P Rockwell, L Witte, C Burrow, M Z Ratajczak, and A M Gewirtz. Stk-1, the human homolog of flk-2/flt-3, is selectively expressed in cd34+ human bone marrow cells and is involved in the proliferation of early progenitor/stem cells. *Proc Natl Acad Sci U S A*, 91(2):459–463, Jan 1994.

[191] C E Carow, E Kim, A L Hawkins, H D Webb, C A Griffin, E W Jabs, C I Civin, and D Small. Localization of the human stem cell tyrosine kinase-1 gene (flt3) to 13q12–¿q13. *Cytogenet Cell Genet*, 70(3-4):255–257, 1995.

[192] N Maroc, R Rottapel, O Rosnet, S Marchetto, C Lavezzi, P Mannoni, D Birnbaum, and P Dubreuil. Biochemical characterization and analysis of the transforming potential of the flt3/flk2 receptor tyrosine kinase. *Oncogene*, 8(4):909–918, Apr 1993.

[193] O Rosnet, H J Bühring, S Marchetto, I Rappold, C Lavagna, D Sainty, C Arnoulet, C Chabannon, L Kanz, C Hannum, and D Birnbaum. Human flt3/flk2 receptor tyrosine kinase is expressed at the surface of normal and malignant hematopoietic cells. *Leukemia*, 10(2):238–248, Feb 1996.

[194] D L Stirewalt and J P Radich. The role of flt3 in haematopoietic malignancies. *Nat Rev Cancer*, 3(9):650–665, Sep 2003.

[195] S D Lyman, L James, J Zappone, P R Sleath, M P Beckmann, and T Bird. Characterization of the protein encoded by the flt3 (flk2) receptor-like tyrosine kinase gene. *Oncogene*, 8(4):815–822, Apr 1993.

[196] C E Carow, M Levenstein, S H Kaufmann, J Chen, S Amin, P Rockwell, L Witte, M J Borowitz, C I Civin, and D Small. Expression of the hematopoietic growth factor receptor flt3 (stk-1/flk2) in human leukemias. *Blood*, 87(3):1089–1096, Feb 1996.

[197] M Walter, I S Lucet, O Patel, S E Broughton, R Bamert, N K Williams, E Fantino, A F Wilks, and J Rossjohn. The 2.7 a crystal structure of the autoinhibited human c-fms kinase domain. *J Mol Biol*, 367(3):839–847, Mar 2007.

[198] C Hannum, J Culpepper, D Campbell, T McClanahan, S Zurawski, J F Bazan, R Kastelein, S Hudak, J Wagner, and J Mattson. Ligand for flt3/flk2 receptor tyrosine kinase regulates growth of haematopoietic stem cells and is encoded by variant rnas. *Nature*, 368(6472):643–648, Apr 1994.

[199] S D Lyman, L James, L Johnson, K Brasel, P de Vries, S S Escobar, H Downey, R R Splett, M P Beckmann, and H J McKenna. Cloning of the human homologue of the murine flt3 ligand: a growth factor for early hematopoietic progenitor cells. *Blood*, 83(10):2795–2801, May 1994.

[200] S D Lyman, L James, S Escobar, H Downey, P de Vries, K Brasel, K Stocking, M P Beckmann, N G Copeland, and L S Cleveland. Identification of soluble and membrane-bound isoforms of the murine flt3 ligand generated by alternative splicing of mrnas. *Oncogene*, 10(1):149–157, Jan 1995.

[201] S D Lyman and S E Jacobsen. c-kit ligand and flt3 ligand: stem/progenitor cell factors with overlapping yet distinct activities. *Blood*, 91(4):1101–1134, Feb 1998.

[202] A M Turner, N L Lin, S Issarachai, S D Lyman, and V C Broudy. Flt3 receptor expression on the surface of normal and malignant human hematopoietic cells. *Blood*, 88(9):3383–3390, Nov 1996.

[203] F Hayakawa, M Towatari, H Kiyoi, M Tanimoto, T Kitamura, H Saito, and T Naoe. Tandem-duplicated flt3 constitutively activates stat5 and map kinase and introduces autonomous cell growth in il-3-dependent cell lines. *Oncogene*, 19(5):624–631, Feb 2000.

[204] C Choudhary, C Müller-Tidow, W E Berdel, and H Serve. Signal transduction of oncogenic flt3. *Int J Hematol*, 82(2):93–99, Aug 2005.

[205] B Hallberg, S I Rayter, and J Downward. Interaction of ras and raf in intact mammalian cells upon extracellular stimulation. *J Biol Chem*, 269(6):3913–3916, Feb 1994.

[206] G L Johnson and R R Vaillancourt. Sequential protein kinase reactions controlling cell growth and differentiation. *Curr Opin Cell Biol*, 6(2):230–238, Apr 1994.

[207] K Mackarehtschian, J D Hardin, K A Moore, S Boast, S P Goff, and I R Lemischka. Targeted disruption of the flk2/flt3 gene leads to deficiencies in primitive hematopoietic progenitors. *Immunity*, 3(1):147–161, Jul 1995.

[208] T Naoe and H Kiyoi. Normal and oncogenic flt3. *Cell Mol Life Sci*, 61(23):2932–2938, Dec 2004.

[209] H Kiyoi, M Towatari, S Yokota, M Hamaguchi, R Ohno, H Saito, and T Naoe. Internal tandem duplication of the flt3 gene is a novel modality of elongation mutation which causes constitutive activation of the product. *Leukemia*, 12(9):1333–1337, Sep 1998.

[210] D L Stirewalt, K J Kopecky, S Meshinchi, F R Appelbaum, M L Slovak, C L Willman, and J P Radich. Flt3, ras, and tp53 mutations in elderly patients with acute myeloid leukemia. *Blood*, 97(11):3589–3595, Jun 2001.

[211] S Meshinchi, W G Woods, D L Stirewalt, D A Sweetser, J D Buckley, T K Tjoa, I D Bernstein, and J P Radich. Prevalence and prognostic significance of flt3 internal tandem duplication in pediatric acute myeloid leukemia. *Blood*, 97(1):89–94, Jan 2001.

[212] C Thiede, C Steudel, B Mohr, M Schaich, U Schäkel, U Platzbecker, M Wermke, M Bornhäuser, M Ritter, A Neubauer, G Ehninger, and T Illmer. Analysis of

flt3-activating mutations in 979 patients with acute myelogenous leukemia: association with fab subtypes and identification of subgroups with poor prognosis. *Blood*, 99(12):4326–4335, Jun 2002.

[213] Y Yamamoto, H Kiyoi, Y Nakano, R Suzuki, Y Kodera, S Miyawaki, N Asou, K Kuriyama, F Yagasaki, C Shimazaki, H Akiyama, K Saito, M Nishimura, T Motoji, K Shinagawa, A Takeshita, H Saito, R Ueda, R Ohno, and T Naoe. Activating mutation of d835 within the activation loop of flt3 in human hematologic malignancies. *Blood*, 97(8):2434–2439, Apr 2001.

[214] A R Leach. *Molecular Modeling principles and applications*. Pearson Prentice hall, second edition edition, 1997.

[215] S B Needleman and C D Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*, 48(3):443–453, Mar 1970.

[216] T F Smith and M S Waterman. Identification of common molecular subsequences. *J Mol Biol*, 147(1):195–197, Mar 1981.

[217] W R Pearson. Rapid and sensitive sequence comparison with fastp and fasta. *Methods Enzymol*, 183:63–98, 1990.

[218] S F Altschul, W Gish, W Miller, E W Myers, and D J Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–410, Oct 1990.

[219] M S Johnson and J P Overington. A structural basis for sequence comparisons. an evaluation of scoring methodologies. *J Mol Biol*, 233(4):716–738, Oct 1993.

[220] A Krogh, M Brown, I S Mian, K Sjölander, and D Haussler. Hidden markov models in computational biology. applications to protein modeling. *J Mol Biol*, 235(5):1501–1531, Feb 1994.

[221] H D Hltie, W Sippl, D Rognan, and G Folkers. *Molecular Modeling*. Wiley-VCH, second edition edition, 2003.

[222] A Sali. Modeling mutations and homologous proteins. *Curr Opin Biotechnol*, 6(4):437–451, Aug 1995.

[223] M C Peitsch and C V Jongeneel. A 3-d model for the cd40 ligand predicts that it is a compact trimer similar to the tumor necrosis factors. *Int Immunol*, 5(2):233–238, Feb 1993.

[224] M Levitt. Accurate modeling of protein conformation by automatic segment matching. *J Mol Biol*, 226(2):507–533, Jul 1992.

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

[226] J M Haile. *Molecular Dynamics Simulation, elementary methods*. Number 1. Wiley-Interscience, 1997.

[227] T Hunter. Signaling–2000 and beyond. *Cell*, 100(1):113–127, Jan 2000.

[228] B J Druker, S Tamura, E Buchdunger, S Ohno, G M Segal, S Fanning, J Zimmermann, and N B Lydon. Effects of a selective inhibitor of the abl tyrosine kinase on the growth of bcr-abl positive cells. *Nat Med*, 2(5):561–566, May 1996.

[229] C L Sawyers, A Hochhaus, E Feldman, J M Goldman, C B Miller, O G Ottmann, C A Schiffer, M Talpaz, F Guilhot, M W Deininger, T Fischer, S G O'Brien, R M Stone, C B Gambacorti-Passerini, N H Russell, J J Reiffers, T C Shea, B Chapuis, S Coutre, S Tura, E Morra, R A Larson, A Saven, C Peschel, A Gratwohl, F Mandelli, M Ben-Am, I Gathmann, R Capdeville, R L Paquette, and B J Druker. Imatinib induces hematologic and cytogenetic responses in patients with chronic myelogenous leukemia in myeloid blast crisis: results of a phase ii study. *Blood*, 99(10):3530–3539, May 2002.

[230] M S Holtz, M L Slovak, F Zhang, C L Sawyers, S J Forman, and R Bhatia. Imatinib mesylate (sti571) inhibits growth of primitive malignant progenitors in chronic myelogenous leukemia through reversal of abnormally increased proliferation. *Blood*, 99(10):3792–3800, May 2002.

[231] P A Harris, M Cheung, R N Hunter, M L Brown, J M Veal, R T Nolte, L Wang, W Liu, R M Crosby, J H Johnson, A H Epperly, R Kumar, D K Luttrell, and J A Stafford. Discovery and evaluation of 2-anilino-5-aryloxazoles as a novel class of vegfr2 kinase inhibitors. *J Med Chem*, 48(5):1610–1619, Mar 2005.

[232] J Stamos, M X Sliwkowski, and C Eigenbrot. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J Biol Chem*, 277(48):46265–46272, Nov 2002.

[233] E R Wood, A T Truesdale, O B McDonald, D Yuan, A Hassell, S H Dickerson, B Ellis, C Pennisi, E Horne, K Lackey, K J Alligood, D W Rusnak, T M Gilmer, and L Shewchuk. A unique structure for epidermal growth factor receptor bound to gw572016 (lapatinib):relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res*, 64(48):6652–6659, Sep 2004.

[234] M Mohammadi, S Froum, J M Hamby, M C Schroeder, R L Panek, G H Lu, A V Eliseenkova, D Green, J Schlessinger, and S R Hubbard. Crystal structure of an angiogenesis inhibitor bound to the fgf receptor tyrosine kinase domain. *EMBO J*, 17(20):5896–5904, Oct 1998.

[235] B Nagar, W G Bornmann, P Pellicena, T Schindler, D R Veach, W T Miller, B Clarkson, and J Kuriyan. Crystal structures of the kinase domain of c-abl in complex with the small molecule inhibitors pd173955 and imatinib (sti-571). *Cancer Res*, 62(15):4236–4243, Aug 2002.

[236] X Zhu, J L Kim, J R Newcomb, P E Rose, D R Stover, L M Toledo, H Zhao, and K A Morgenstern. Structural analysis of the lymphocyte-specific kinase lck in complex with non-selective and src family selective kinase inhibitors. *Structure*, 7(6):651–661, Jun 1999.

[237] F Meggio, A Donella Deana, M Ruzzene, A M Brunati, L Cesaro, B Guerra, T Meyer, H Mett, D Fabbro, and P Furet. Different susceptibility of protein kinases to staurosporine inhibition. kinetic studies and molecular bases for the resistance of protein kinase ck2. *Eur J Biochem*, 234(1):317–322, Nov 1995.

[238] N Schiering, S Knapp, M Marconi, M M Flocco, J Cui, R Perego, L Rusconi, and C Cristiani. Crystal structure of the tyrosine kinase domain of the hepatocyte growth factor receptor c-met and its complex with the microbial alkaloid k-252a. *Proc Natl Acad Sci U S A*, 100(22):12654–12659, Oct 2003.

[239] S Atwell, J M Adams, J Badger, M D Buchanan, I K Feil, K J Froning, X Gao, J Hendle, K Keegan, B C Leon, H J Müller-Dieckmann, V L Nienaber, B W Noland, K Post, K R Rajashankar, A Ramos, M Russell, S K Burley, and S G Buchanan. A novel mode of gleevec binding is revealed by the structure of spleen tyrosine kinase. *J Biol Chem*, 279(53):55827–55832, Dec 2004.

[240] M Mohammadi, G McMahon, L Sun, C Tang, P Hirth, B K Yeh, S R Hubbard, and J Schlessinger. Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. *Science*, 276(5314):955–960, May 1997.

[241] C A Pickover. Spectrographic representation of globular protein breathing motions. *Science*, 223(4632):181–182, Jan 1984.

[242] A Bairoch. The enzyme database in 2000. *Nucleic Acids Res*, 28(1):304–305, Jan 2000.

[243] J D Thompson, T J Gibson, F Plewniak, F Jeanmougin, and D G Higgins. The clustal_x windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, 25(24):4876–4882, Dec 1997.

[244] G H Gonnet, M A Cohen, and S A Benner. Exhaustive matching of the entire protein sequence database. *Science*, 256(5062):1443–1445, Jun 1992.

[245] S M Bishop, J B Ross, and R A Kohanski. Autophosphorylation dependent destabilization of the insulin receptor kinase domain: tryptophan-1175 reports changes in the catalytic cleft. *Biochemistry*, 38(10):3079–3089, Mar 1999.

[246] P M Chan, S Ilangumaran, J La Rose, A Chakrabartty, and R Rottapel. Autoinhibition of the kit receptor tyrosine kinase by the cytosolic juxtamembrane region. *Mol Cell Biol*, 23(9):3067–3078, May 2003.

[247] J C Williams, A Weijland, S Gonfloni, A Thompson, S A Courtneidge, G Superti-Furga, and R K Wierenga. The 2.35 a crystal structure of the inactivated form of chicken src: a dynamic molecule with multiple regulatory interactions. *J Mol Biol*, 274(5):757–775, Dec 1997.

[248] X Liu, S R Brodeur, G Gish, Z Songyang, L C Cantley, A P Laudano, and T Pawson. Regulation of c-src tyrosine kinase activity by the src sh2 domain. *Oncogene*, 8(5):1119–1126, May 1993.

[249] Tripos Inc. Sybyl 7.3, 1699 South Hanley Road, St. Louis, MO, 63144.

[250] C D Mol, K B Lim, V Sridhar, H Zou, E Y Chien, B C Sang, J Nowakowski, D B Kassel, C N Cronin, and D E McRee. Structure of a c-kit product complex reveals the basis for kinase transactivation. *J Biol Chem*, 278(34):31461–31464, Aug 2003.

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

[252] H J C Berendsen, D van der Spoel, and R van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, Sep 1995.

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

[254] D Van Der Spoel, E Lindahl, B Hess, G Groenhof, A E Mark, and H J Berendsen. Gromacs: fast, flexible, and free. *J Comput Chem*, 26(16):1701–1718, Dec 2005.

[255] J Hermans, W F Berendsen, H J C Van Gunsteren, and J P M Postma. consistent empirical potential for water-protein interactions. *Biopolymers*, 23(8):1513–1518, 1984.

[256] G A Kaminski, R A Friesner, J Tirado-Rives, and W L Jorgensen. Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. *Journal of Physical Chemistry B*, 105(28):6474–6487, 2001.

[257] U Essmann, L Perera, M L Berkowitz, T Darden, H Lee, and L G Pedersen. A smooth particle mesh ewald method. *Journal of Physical Chemistry*, 103(19):8577–8593, Nov 1995.

[258] 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. *The Journal of Chemical Physics*, 81(8):3684–3690, 1984.

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

[260] W Humphrey, A Dalke, and Schulten K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[261] B Speelman, B R Brooks, and C B Post. Molecular dynamics simulations of human rhinovirus and an antiviral compound. *Biophys J*, 80(1):121–129, Jan 2001.

[262] B J Druker. Sti571 (gleevec) as a paradigm for cancer therapy. *Trends Mol Med*, 8(4 Suppl):14–18, 2002.

[263] M A Arslan, O Kutuk, and H Basaga. Protein kinases as drug targets in cancer. *Curr Cancer Drug Targets*, 6(7):623–634, Nov 2006.

[264] J D Watson and F H Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, Apr 1953.

[265] K A Thiel. Structure-aided drug design's next generation. *Nat Biotechnol*, 22(5):513–519, May 2004.

[266] M A Marti-Renom, M S Madhusudhan, A Fiser, B Rost, and A Sali. Reliability of assessment of protein structure prediction methods. *Structure*, 10(3):435–440, Mar 2002.

[267] A M Lesk and C Chothia. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J Mol Biol*, 136(3):225–270, Jan 1980.

[268] C Chothia and A M Lesk. The relation between the divergence of sequence and structure in proteins. *EMBO J*, 5(4):823–826, Apr 1986.

[269] J H Till, A J Ablooglu, M Frankel, S M Bishop, R A Kohanski, and S R Hubbard. Crystallographic and solution studies of an activation loop mutant of the insulin receptor tyrosine kinase: insights into kinase mechanism. *J Biol Chem*, 276(13):10049–10055, Mar 2001.

[270] H Chen, J Ma, W Li, A V Eliseenkova, C Xu, T A Neubert, W T Miller, and M Mohammadi. A molecular brake in the kinase hinge region regulates the activity of receptor tyrosine kinases. *Mol Cell*, 27(5):717–730, Sep 2007.

Riassunto

L'argomento di questa tesi è la parte in comune delle proteine appartenenti alla famiglia delle tirosine chinasi (TKs): il dominio kinasico (KD). Questa è la porzione enzimatica che catalizza il trasferimento di un gruppo fosforico da una molecola di ATP ad un residuo di tirosina di un substrato peptidico.

Le cellule hanno sviluppato questo espediente attraverso molti processi evolutivi fino al raggiungimento di una stabiltà nel funzionamento e nel controllo negli organismi eucarioti multicellulari. Attraverso il rimescolamento e la duplicazione genica, questa famiglia di proteine si è allargata e differenziata in membri con topologie differenti. La gamma di diverse subunità proteiche che si trovano in posizione N- e C-terminale al KD caratterizza le TKs secondo le loro funzioni e comportamenti biologici, le interazioni con partners secondari e la localizzazione nella cellula. Due gruppi di protein tirosin chinasi, transmembrana e citosoliche, sono presenti nella cellula e prendono parte allo scambio di diversi tipi di segnali. I membri del primo gruppo sono capaci di captare un segnale specifico dall'ambiente extracellulare e di transdurlo all'interno della cellula; vengono chiamati recettori tirosin chinasici. Le proteine del secondo gruppo vengono spesso ritrovate in vie biologiche dove trasferiscono e differenziano segnali diversi e vengono chiamate tirosin chinasi non recettoriali. Nell'uomo, 90 tirosin chinasi sono state identificate e classificate in 30 sottofamiglie per le loro similitudini nelle sequenze aminoacidiche.

Il ruolo cruciale di queste proteine nella crescita, sviluppo e differenziazione cellulare è stato confermato da numerosi articoli scientifici in letteratura così come la loro partecipazione in numerose malattie. Infatti, diverse modificazioni proteiche come mutazioni, delezioni e inserzioni di uno o più aminoacidi sono riscontrate in diverse patologie dove l'attività enzimatica delle TKs risulta anormale. Inoltre, sovraespressioni e fusioni di parte di queste proteine sono state isolate da linee cellulari oncogeniche. Perciò la necessità di comprendere il meccanismo e la possibilità di trattare molte malattie dove sono implicate le TKs, ha spinto la comunità scientifica a studiare vari aspetti di queste proteine.

Il lavoro qui presentato va in questa direzione. La sequenza, la struttura, la dinamica e l'interazione col ligando sono alcune caratteristiche del dominio chinasico prese in considerazione in questo studio. Questioni riguardanti il riconoscimento del ligando, le caratteristiche strutturali determinate da variazioni nella sequenza e la dinamica dell'attivazione dell'enzima sono state poste al fine di esplorare il dominio chinasico a livello molecolare attraverso diversi strumenti computazionali.

Lo studio incomincia con lo screening di diverse banche dati al fine di ottenere informazioni riguardanti il KD. Le sequenze aminoacidiche delle TKs sono state analizzate e studiate attraverso l'allineamenti multipli per valutare la conservazione dei residui nei vari membri della famiglia. Quindi, le strutture risolte sperimentalmente sono state cercate nel protein databank (PDB) e raccolte.

I complessi ligando-KD sono stati comparati per trovare i requisiti minimi necessari a una molecola per interagire nel sito di legame dell'ATP. Con la definizione "ligand-core", nel secondo capitolo, vengono identificati gli elementi strutturali essenziali in un composto per potersi legare alle TKs. Questa parte del ligando interagisce con il dominio catalitico attraverso un legame idrogeno e diversi contatti di van der Waals con residui il cui carattere idrofobico è conservato nei diversi membri della famiglia.

Successivamente, sono state messe a confronto le strutture cristallografiche delle KDs per evidenziare le parti tridimensionali conservate così come le deviazioni conformazionali dovute a differenze nella struttura primaria. Nel terzo capitolo, cinque gruppi (clusters) di aminoacidi sono stati identificati nel KD e sembrano avere un ruolo principale nella conformazione di questi enzimi. In base alla disposizione di questi clusters nei diversi stati enzimatici, è stata proposta una classificazione di queste proteine nel tentativo di accumunare gli elementi con dinamica molecolare simile. In particolare, 14 residui all'interfaccia tra il lobo N e il lobo C sembrano essere di cruciale importanza per la conformazione strutturale ed il comportamento dinamico di queste proteine. Il tipo di aminoacidi e le loro interazioni (2 polari e molti contatti idrofobici) risultano conservati nelle sottofamiglie delle TKs. In questo capitolo, sono inoltre descritte altre caratteristiche dei ligandi delle tirosin chinasi, come la conformazione, le interazioni e la farmacologia. Infine, sono riportati esempi di riconoscimento ligando-proteina; nel caso delle tirosin chinasi vengono riscontrati due meccanismi: i) l'adattamento della proteina al ligando e ii) la conformazione proteica complementare a quella del ligando (teoria chiave-serratura).

Il capitolo quarto comincia con la validazione delle simulazioni dinamiche molecolari (molecular dynamics simulation, MD) come metodo per studiare i movimenti del dominio chinasico. Diverse traiettorie sono state calcolate, via MD, per la conformazione non-produttiva e produttiva del recettore insulinico (Insr) al fine di capire la riproducibilità di questo tipo di simulazioni. Il cuore di questo capitolo è dedicato all'esplorazione di un possibile meccanismo d'attivazione enzimatica riguardante il dominio chinasico.

La proteina Flt3, coinvolta nella normale ematopoiesi, è implicata in alcune malattie come la leucemia mieloide acuta (AML) quando la sua attività enzimatica è fuori controllo per la presenza di una doppia duplicazione interna (interna tandem duplication, ITD) a livello del dominio juxtamembrana (JM). La funzione autoinibitoria del JM nei riguardi del dominio chinasico è stata rivelata dalla struttura cristallografica; l'anormale inserzione di ITDs nello juxtamembrana, porta all'attivazione costitutiva dell'enzima. I movimenti del dominio catalitico di Flt3 sono stati predetti via MD sia per la struttura della proteina nella sua forma wild-type che per diverse forme aventi ITDs. Quest'ultime strutture sono state generate virtualmente grazie a tecniche di modellazione per omologia (homology modeling). Con queste prime valutazioni, il meccanismo inibitorio del JM è stato interpretato come ostacolo sterico agente sull'elica C, ed è stato successivamente approfondito con lo studio di simulazioni su costrutti di Flt3 dove parti dello juxtamembrana sono stati rimossi. La libertà di movimento dell'elica C ottenuta dalla rimozione dell'interazione, anche parziale, del dominio autoinibitore, è stata identificata come la forza motrice per i movimenti del resto del lobo N.

Gli aminoacidi all'interfaccia tra i due lobi, identificati nel terzo capitolo, sono stati ulteriormente esaminati nel quinto. Le dinamiche molecolari delle strutture cristallografiche non-produttiva e produttiva di Insr così come una terza conformazione (intermedia tra i due stati) sono state studiate ponendo particolare attenzione sui residui all'interfac-

cia. Mentre la conformazione della maggioranza di questi residui resta stabile durante le simulazioni, un riarrangiamento dell'interazione per una triade di aminoacidi in prossimità dell'elica C è visibile. Ulteriori cambiamenti conformazionali sono apprezzabili per le catene laterali di due residui, uno della "hinge region" e il secondo del loop in posizione C-terminale rispetto all'elica C.

Nello sviluppo di ligandi basato sulla struttura del bersaglio, una delle principali questioni, è la scelta della migliore conformazione proteica come candidato molecolare. Spesso, le strutture sperimentali disponibili non sono sufficienti a colmare le necessità ed in alternativa la tecnica *in-silico* "homology modeling" può essere usata per predire la struttura proteica desiderata. Questo metodo è basato sull'identificazione delle regioni strutturalmente conservate (SCRs) nella proteina templato da trasporre alla sequenza obiettivo. L'isolamento di SCRs dovrebbe essere il più accurato possibile per poter generare un modello tridimensionale stabile e significativo. In questo lavoro, sono state valutate le strutture di Insr come possibili templati generali per l'homology modeling delle tirosin chinasi. Inoltre, sono state proposte come SCRs i residui all'interfaccia tra i due lobi.

Concludendo, l'identificazione del "ligand-core" può essere usato come farmacoforo minimo nel caso in cui le interazioni KD-ligando siano analizzate come per esempio nei protocolli di "virtual screening". 14 residui sono stati isolati all'interfaccia tra i due lobi del KD e le loro normali fluttuazioni dinamiche sono state valutate per tre differenti conformazioni del recettore per l'insulina. Di fatto, questi 14 residui sono buoni candidati tra le possibili sequenze aminoacidiche implicate nella stabilizzazione delle conformazioni diversamente popolate del dominio chinasico. Infine, il meccanismo di attivazione della proteina Flt3 nella forma wild-type ed oncogenica è stato esaminato. Certi elementi del KD, in particolare l'elica C, sono stati proposti come componenti principali durante il cambiamento conformazionale che porta all'attivazione enzimatica.

Résumé

Le domaine commun des protéines tyrosine kinases (TKs), le domaine kinase (DK), est l'objet de la présente thèse. Il constitue l'entité enzymatique catalysant le transfert d'un groupement phosphate d'une molécule d'ATP à un résidu tyrosine du substrat peptidique. La structure primaire du domaine catalytique rassemble environ 300 acides aminés qui se replient pour former une protéine soluble d'environ 30 KDa localisée dans le cytoplasme.

Dans la figure 6.1, l'on aperçoit sa structure bilobée classique, c'est-à-dire le lobe N-terminal et le lobe C-terminal connectés par un segment nommé "hinge region".
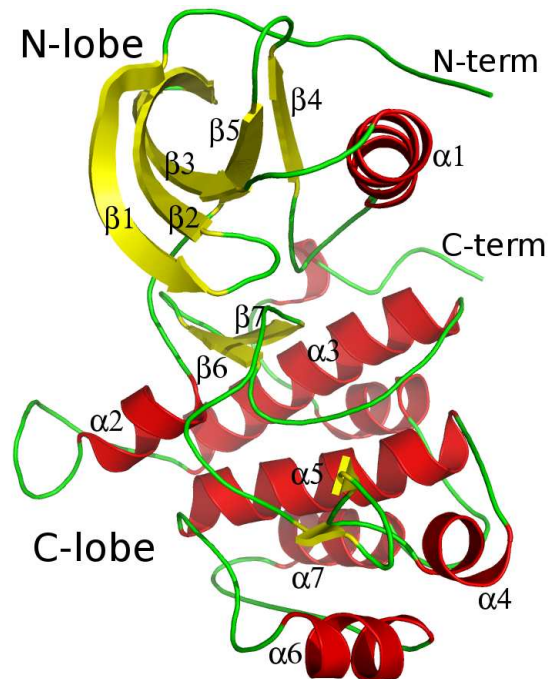
Fig. 6.1: *L'arrangement en 3D du domaine catalytique. Les éléments de structure secondaire de la structure cristalline du récepteur de l'insuline (abrégé Insr, code PDB : 1IRK) sont représentés en "cartoon" et colorées comme suit : jaune pour les feuillets, vert pour les boucles et rouge pour les hélices. Les coordonnées atomiques ont été déssinées en utilisant le logiciel* PYMOL *[1].*

La terminaison aminée du domaine formant le lobe N-terminal comporte une partie centrale majeure composée de feuillets β antiparallèles et d'une hélice α (α1) (voir image 6.1), appelé hélice C enjambant les feuillets β 3 et 4. Le lobe C-terminal a un noyau hélicoïdal de six hélices α conservées et un petit feuillet β (voir image 6.1).

D'après les travaux accumulés sur la protéine [2], ses fonctions biophysiques/biochimiques

principales sont les suivantes :

i) La liaison et l'orientation des cofacteurs ou contre-ions du complexe ; principalement l'ATP, mais aussi le GTP, avec des phosphates coordonnés par un ou deux cations divalents Mg^{++} ou Mn^{++} ; ii) La liaison et l'orientation du substrat peptidique ; iii) Le transfert du phosphate γ du cofacteur à un hydroxyl d'une chaîne latérale d'un résidu (soit la sérine, la thréonine ou la tyrosine) du substrat.

Les cellules ont développé ce système basé sur les protéines kinases durant de multiples étapes de l'évolution jusqu'à atteindre une fonction et un contrôle stables dans les eukaryotes multicellulaires [3]. Par le biais de réagencements de domaines et de dédoublement de gènes, la famille de protéines en question s'est élargie et s'est différenciée en divers éléments ayant différentes topologies [4]. C'est la variation de modules protéiques des côtés N- et C-terminaux du DK qui détermine les fonctions biologiques, les comportements, les interactions avec des partenaires secondaires et la localisation cellulaire [5]. Il existe deux groupes de kinases : les protéines transmembranaires et les protéines cytosoliques, les deux étant impliquées dans la signalisation cellulaire [6]. Les premières, appelées "recepteur tyrosine kinases", sont capables de détecter un signal spécifique de l'environnement extracellulaire et de le transmettre dans la cellule. Cependant, les dernières, nommées "non-recepteur tyrosine kinases", sont souvent trouvées dans les voies de signalisation biologiques où ils transfèrent et modulent divers signaux dans la cellule. Chez l'homme, 90 tyrosine kinases ont été identifiées et classifiées en 30 sous-familles par similarité de séquence [7].

Dans des conditions physiologiques, le rôle principal des protéines tyrosine kinases est la croissance, la différentiation et la communication cellulaire [6]. Dans des conditions pathologiques par contre, la dérégulation de ces enzymes a principalement été corrélée à l'apparition de tumeurs, du diabète et d'autres pathologies [8].

Beaucoup de modifications protéiques comme les mutations, les délétions et les insertions de quelques d'acides aminés expliquent le fonctionnement anormal de plusieurs TKs [8]. De plus, des phénomènes comme la surexpression et la fusion de protéines ont été identifiés dans des lignées cellulaires à comportement oncogénique. Pour ces raisons, la necessité de comprendre les mécanismes des maladies liées aux TKs contraint la communauté scientifique à analyser plusieurs aspects de ces protéines et à chercher un contrôle extérieur approprié.

Les inhibiteurs des TKs les mieux connus et les plus efficaces cliniquement miment principalement l'adénine, ce sont donc des molécules entrant en compétition avec l'ATP se liant à la même poche de la protéine [9].

Une grande découverte à été faite avec l'Imatinib, ayant de la sélectivité contre la protéine de fusion oncogénique Bcr-Abl, et les protéines Kit et Pdgfr [10]. En 2001, cette molécule a obtenu l'autorisation de mise sur le marché de la part de la FDA comme traitement contre la leucémie myéloïde chronique (LMC) [10].

La présente thèse évalue plusieurs aspects de ces entités enzymatiques proches sur le plan évolutif en mettant l'accent sur les séquences, les structures, les dynamiques et les liaisons de ligands respectives. Afin de saisir le domaine de kinase au niveau moléculaire, des thèmes comme la reconnaissance moléculaire, les points communs de la liaison d'un ligand ou les bouts d'acides aminés déterminant les caractéristiques structurales et la dynamique de l'activation enzymatique sont abordés et explorés par des outils informatiques.

Comme première étape, des bases de données ont été criblées pour rassembler des informations sur le KD. Dans le but d'évaluer la conservation d'acides aminés parmi les membres de la famille, les séquences de 90 tyrosine kinases ont été obtenues de la base de données EXPASY (at us.expasy.org/sprot/) et alignées. Toutefois, les 138 structures disponibles du domaine catalytique expérimentalement déterminées représentant 30 TKs uniques ont été téléchargées du site web de la "protein databank" (PDB) (at www.rcsb.org/pdb/home/home.do) et regardées de près.

Le deuxiéme chapitre commence avec une comparaison des complexes DK-ligand afin de trouver des caractéristiques de liaison communes. Pour chaque complexe, les résidus intéragissant avec le ligand ont été étudiés en visualisant la région protéique se situant 4.5 Å autour du ligand. Les acides aminés identifiés ont été comparés au niveau de la séquence par alignement de séquence multiple de toutes les structures considérées. Cette procédure a permis de mettre en évidence les résidus communs impliqués dans des interactions conservées entre les ligands et les DKs. Dans tous les complexes, ce type commun d'interactions consiste en des interactions hydrophobes incluant quatre résidus et un pont hydrogène unique avec un atome d'azote du squelette; ceci est visualisé pour le récepteur de l'insuline (Insr) dans l'image 6.2.
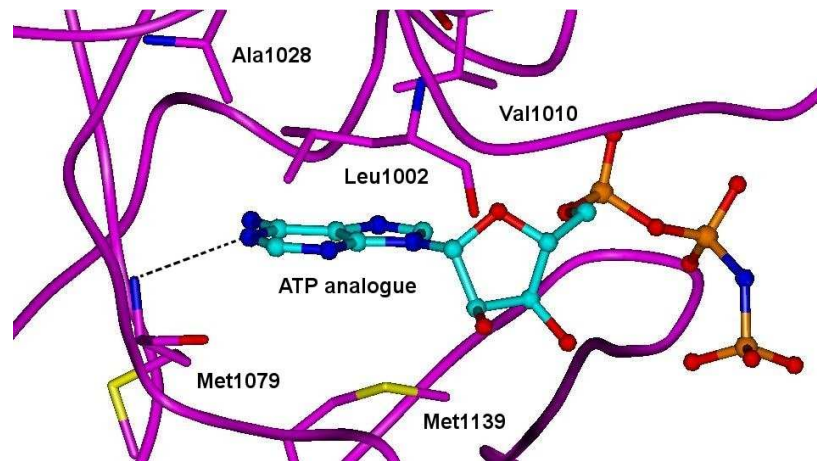
FIG. 6.2: *Les types d'interactions communs dans les complexes ligand-DK. La structure secondaire du DK du récepteur de l'insuline (1IR3) est représentée en tube (magenta), l'analogue de l'ATP apparaît en "sphères et bâtonnets" et les résidus communs en bâtonnets, selon un code de couleurs (le carbone en bleu clair pour le ligand et en magenta pour la protéine, l'oxygène en rouge, l'azote en bleu foncé, le soufre en jaune, le phosphore en orange et les ponts hydrogène en pointillés noirs).*

Pour les chaînes latérales des 4 résidus prenant part à des liasons van der Waals avec les ligands, un consensus concernant le caractère hydrophobe est trouvé partout dans les séquences de la famille des TKs (voir tableau 6.1). À considéérer le ligand, les caractéristiques minimales (une moitié plane et hydrophobe contenante un atome accepteur de liason hydrogène) pour qu'une molécule interagisse avec le site de liaison de l'ATP sont appelée "noyau du ligand".

Puis, les structures cristallines disponibles ont été comparées afin de comprendre la conservation tridimensionnelle ainsi que les variations dues aux changements au niveau de la structure primaire. Les différences dans l'arrangement des portions des DKs ont abouti

TAB. 6.1: *Variabilité des acides aminés du type hydrophobe intéragissant avec le noyau du ligand. Variance (en pourcentage) du type d'acide aminé aux positions (numérotation selon le récepteur de l'insuline) impliqué dans les contacts hydrophobes avec le noyau du ligand.*

Leu1002	69%	Leu
	30%	Ile
	1%	Val
Val1010	96%	Val
	3%	Ile
	1%	Leu
Ala1028	92%	Ala
	4%	Ile & Val
	4%	Cys,Phe & Leu
Met1139	89%	Leu
	9%	Met
	2%	Val & Phe

à la définition de cinq groupes hydrophobes (clusters) différents formés par d'acides aminés participant à des mouvements ou fonctions spécifiques : le noyau hydrophobe du lobe N-terminal (H1), celui du lobe C-terminal (H2), le noyau d'interaction (la partie protéique complémentaire au noyau du ligand), la phenylalanine du motif DFG, et une partie de l'hélix C. Dans l'image 6.3, un exemple pour l'état non-productif et productif du Insr est donné.

De plus, une classification des DKs a été proposée, basée sur la disposition observée de ces clusters dans différents états conformationels des DKs. Les réarrangements semblent être caractéristiques pour quatre familles, c'est-à-dire les "Insr-like", les "Cdk-like", les "Src-like" et ceux qui sont probablement "always-active".

L'interface entre les clusters qui appartiennent au noyau hydrophobe des deux lobes prénte 14 résidus assez conservés qui interagissent par deux contacts polaires et par de larges contacts apolaires (voir image 6.4). La configuration particulière de cette région joue un rôle potentialment important dans la disposition et la dynamique du DK.

D'un côté, dans la partie intérieure de la protéine, quelques-uns de ces résidus consituent deux faces opposées dont la fonction pourrait être de permettre les mouvements indépendants des deux lobes. De l'autre côté, la partie extérieure de l'interface a montré deux dents du lobe C-terminal qui encadrent le lobe N-terminal par de residus de la boucle située en C-terminal de la hélix C.

De surcroît, quand les deux structures de l'Insr (1IRK et 1IR3) sont comparés, deux variations du type d'interaction dans l'interface définie sont repérées. La première concerne une rotation d'une liaison simple à la chaîne latérale de la Met1104 appartenant à la "hinge region" entraînant une perte mineure des liasons van der Waals avec la Val1167 et la Lys1174 (chaîne carbonée). La seconde, plus importante, concerne le réagencement entre

(a) *L'Insr non-productif*   (b) *L'Insr productif*
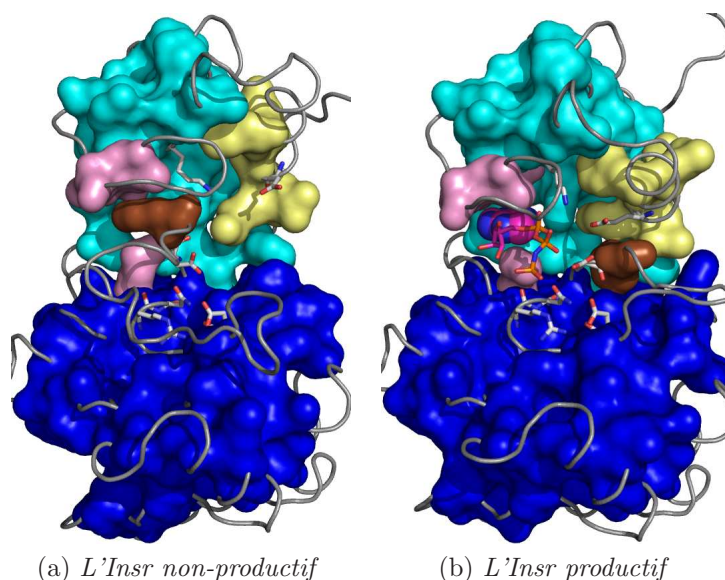
FIG. 6.3: *Représentation des états non-productif (code PDB : 1IRK) et productif (1IR3) de l'Insr tyrosine kinase. Les chaînes protéiques sont représentées comme rubans gris. Les résidus des 5 clusters hydrophobes sont rendus visibles par leur surface moléculaire et par un code de couleurs : en bleu clair pour le cluster 1, en bleu foncé pour le cluster 2, en rose pour le cluster 3, en jaune pour le cluster 4 et en brun pour le cluster 5. Les résidus participant à la catalyse enzymatique comme les ligands sont déssinés en bâtonnets. Le noyau du ligand est visible en CPK. La couleur des atomes est codée de la manière suivante : l'azote en bleu, l'oxygène en rouge, le chlore en vert, le phosphore en orange et le carbone en gris pour les atomes protéiques ; les atomes des ligands sont en magenta.*
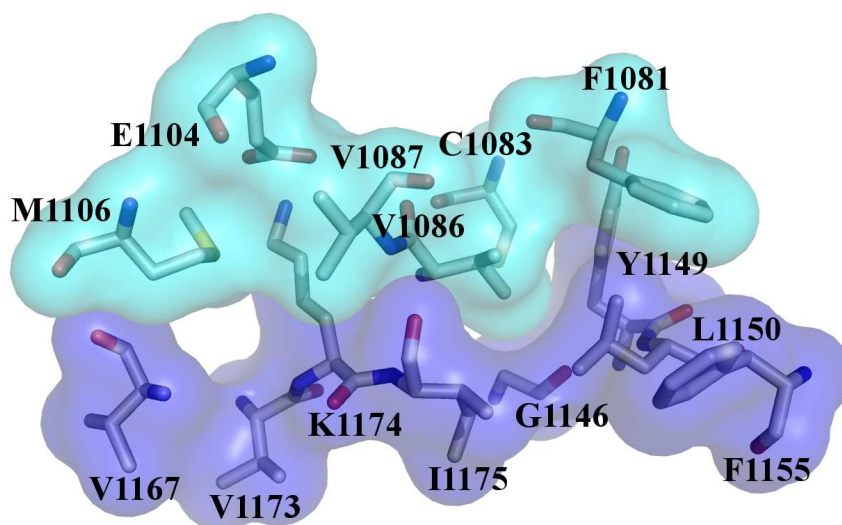
FIG. 6.4: *L'interface entre H1 et H2 pour l'Insr non-productif. Les residus situés à l'interface dans la structure cristalline sont représentés en bâtonnets, étiquetés et chaque type d'atome apparaît dans une couleur bien déterminée : gris pour le carbone, rouge pour l'oxygène, bleu foncé pour l'azote et jaune pour le soufre. Les surfaces des atomes sont aussi visibles et codées en couleur pour les clusters : bleu clair pour H1 et bleu foncé pour H2.*

les résidus Phe1081, Cys1083, Val1086, Leu1150, Phe1155 et Ile1175, avec un réarrangement majeur des résidus numéro 1081, 1150 et 1155. Le changement de conformation pour le dernier résidu est vu fréquemment, même si ce n'est pas toujours le cas, entre les états productifs et non-productifs cristalisés des TKs.

Dans le même chapitre, différents modes de liaison du ligand sont présentés, leurs charactéristiques chimiques et des aspects pharmacologiques sont discutés. En outre, la question de la sélection conformationnelle comparée à la conformation induite par la liaison du ligand est abordée dans le contexte des inhibiteurs des TKs.

Le quatrième chapitre commence par l'évaluation de la simulation de dynamique moléculaire (DM) en tant que moyen d'analyser le mouvement du DK. Plusieurs trajectoires ont été calculées pour les conformations non-productive et productive du récepteur de l'insuline afin de résoudre le problème de la réproductivité des trajectoires. Les systèmes et les protocoles de DM ont été testés en utilisant différentes valeurs de départ pour la génération aléatoire des vitesses initiales. Les trajectoires de chaque état diffèrent. L'intervalle de racine de l'écart-type moyen est plus petit pour les valeurs des paramètres de l'état productif, par rapport à celles de l'état non-productif. La différence de conformation exprimée en racine de l'écart-type moyen est significative si elle dépasse 6% pour l'état productif et plus que 13% pour l'état non-productif.

La partie centrale de ce chapitre est dédiée à l'exploration d'un possible mécanisme d'activation du DK. La protéine Flt3 comportant des duplications en tandem internes (DTIs) au niveau de la juxtamembrane (JM) est impliquée dans des tumeurs malignes hématopoiétiques, car l'insertion de la séquence entraîne l'activation constitutive du DK. Depuis 2004, la forme inhibée de la protéine est connue : le DK formant un complexe serré avec la petite JM [11]. De l'information structurale, il a été déduit que l'activation oncogénique par les DTIs résulte d'une rupture de l'interaction entre une partie de la JM et le DK [11].

Le mouvement de Flt3 a été calculé pour la protéine "wild-type" et pour plusieurs modèles d'homologie contenant des DTIs. L'idée de cette première évaluation par DM, une gêne stérique de la JM agissant sur l'hélix C, a été étudiée plus tard avec des Flt3 où des parties de la JM ont été éliminées. Le premier modèle, Flt3-593, contient la partie de la JM couvrant l'hélix C, mais manque le feuillet β et la région enfouie. Pour le deuxième, Flt3-603, toute la JM manque. Les dynamiques respectives de ces deux systèmes et de la protéine "wild-type" ont été calculées durant 10 ns de la vie protéique par DM.

La protéine Flt3-603 répond immédiatement (avant la première nanoseconde de simulation) au retrait de la JM entière avec un changement majeur principalement du lobe N-terminal. Cependant, la délétion partielle de la JM de Flt3-593 engendre une variation structurale similaire, mais apparaîssant plus tardivement (après la première nanoseconde). Dans la figure 6.5 la nature de ces mouvements est représentée par la superposition de deux conformations de Flt3-603 reproduites à des moments différents. En accord avec les vérifications visuelles, l'analyse différentielle de la racine de l'écart-type moyen montre qu'il y a des régions subissant plus de mouvements que d'autres tandis que le repliement général des trois systèmes reste conservé (voir image 6.6). Les valeurs de la racine de l'écart-type moyen du DK sont relativement bas, allant de 0.8 à 2.7 Å et plutôt stables au cours du temps, tandis qu'il n'y a pas de différences majeures pour le lobe C-terminal ayant des valeurs de racine de l'écart-type moyen sous 1.2 Å pour les trois protéines. Néanmoins, pour le lobe N-terminal, plusiers formes stables et cavantage de mouvement
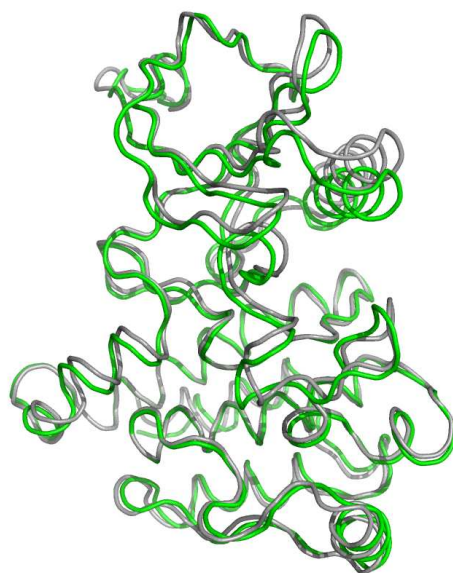
154

FIG. 6.5: *Changements conformationnels de Flt3-603. Les structures secondaires des protéines sont montrées en rubans. Les conformations de Flt3-603 sont colorées selon le temps : En gris pour 0 ns et en vert pour 1 ns.*
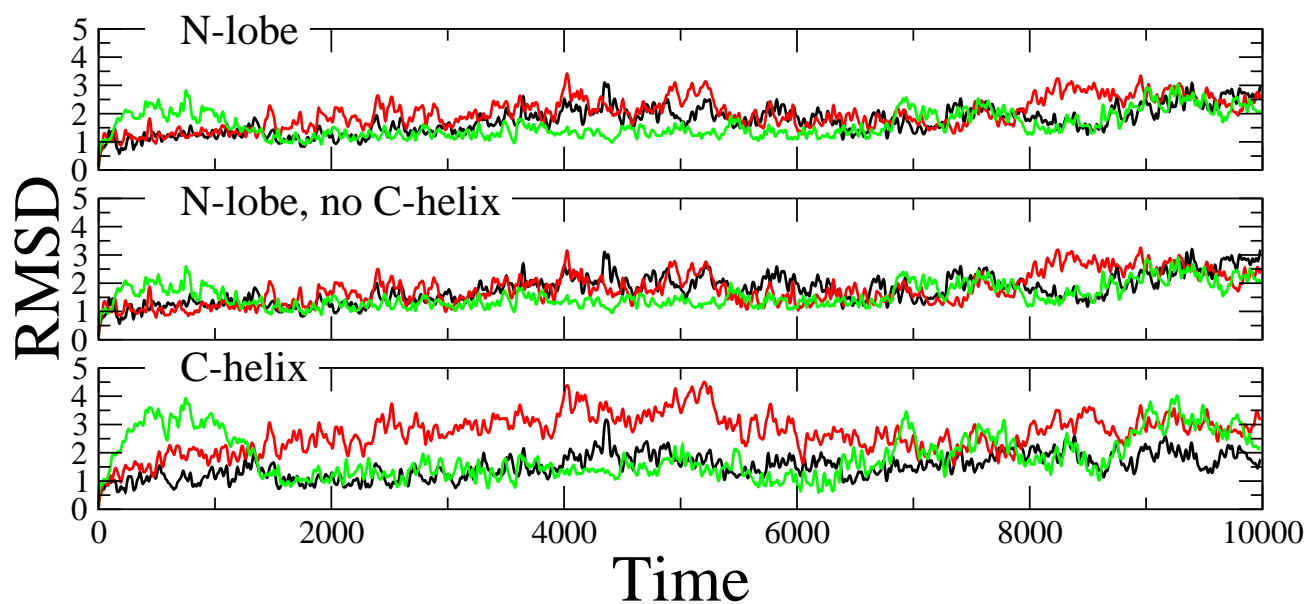

FIG. 6.6: *La stabilité des différentes parties de Flt3. La racine de l'écart-type moyen indiquée en Å exprimée comme fonction du temps en ps pour les atomes Cα du DK. Flt3-wt est coloré en noir, Flt3-593 en rouge et Flt3-603 en vert. Les racines de l'écart-type moyen de différentes portions sont montrées dans leurs graphiques correspondants.*

ont été observés dans les trois systèmes, dont le "wild-type" qui s'est avéré être le plus stable. Une analyse détaillée a clairement démontré que l'hélice C est l'élément structural ayant les valeurs de racine de l'écart-type moyen les plus élevées et présentant les différences les plus significatives des trois systèmes. Pour Flt3-wt, les valeurs de la racine de l'écart-type moyen sont dans l'ordre de 0.6 à 3 Å ; cependant, pour les deux autres ayant une JM plus courte, les valeurs sont significativement plus importantes et varient entre 0.6 et 4.5 Å. Concernant la tendance générale, Flt3-603 a deux pics correspondant au relâchement de la JM entière, le premier dans les premières nanosecondes et le deuxième dans les dernières de la simulation. Par contre, Flt3-593 dispose d'une augmentation plus lente de la racine de l'écart-type moyen, mais d'une persistance plus longue comparée au wild-type. L'amplitude plus importante de la racine de l'écart-type moyen de l'hélice C isolée par rapport au lobe N-terminal (ce dernier élément ayant été soustrait) pourraît signifier que cette hélice est le point de départ de tous les mouvements de la région N-terminale.

En résumant, les relâchements des portions de la JM ont indiqué un changement conformationnel majeur du lobe N-terminal et son hélice C paraît contrôler ce mouvement. Sûrement, des études plus approfondies sont nécessaires pour préciser, prouver ou réfuter ces résultats. Pour qu'il y ait activation, des changements plus importants que ceux du DK doivent se faire, notamment le reagencement du motif DFG et l'ouverture de la boucle d'activation. Quoi qu'il en soit, les résultats présentés suggèrent un rôle majeur de l'hélice C durant la premiére étape de l'activation du DK de la protéine Flt3.

Les résidus analysés au troisième chapitre sont de nouveau investigués au cinquième. Les dynamiques des deux structures crystallines de l'Insr, la non-productive et la productive, ainsi que celle d'une troisième conformation ont été regardées de près en mettant l'accent sur les résidus de l'interface. Parmi les trajectoires calculées, la conformation intermédiaire atteint l'équilibre dynamique avec une déviation initiale (dans la première nanoseconde) plus grande comparée aux autres structures. De manière générale, les trois structures ont été stables avec une rotation mineure du lobe N-terminal par rapport au lobe C-terminal, déterminant ainsi la conformation globale.

De même, la majorité des 14 résidus de l'interface ont été stables pour les trois protéines durant la DM (voir image 6.7). Un réarrangement majeur des interactions pour une triade de résidus (la Phe1081, la Leu1150 et la Phe1155) dans la région C-terminale de la hélice C est apparent pour l'état non-productif. De plus, des variations mineures sont montrés pour les états non-productif et intermédiaire : l'une impliquant la chaîne latérale de la Met1106, appartenant à la "hinge region", et une autre deuxième impliquant la Cys1083 de la boucle C-terminale de l'hélice C.

Durant la DM des deux états mentionnés, la Cys1083 place son atome de soufre de la chaîne latérale entre la Tyr1149 et la Val1086, ce qui éloigne ces résidus l'un de l'autre. En général, les résidus à l'interface de H1 et H2 montrent des fluctuations de leurs atomes de la chaîne latérale de moins de 2 Å à l'exception de la Phe1155 (voir image 6.8). Les interactions polaires impliquant des atomes de la chaîne principale, notamment l'oxygène de Ile1175 et l'azote de Val1087, et des atomes de chaînes latérales, soit le groupement amine de Lys1174 et le carboxyle de Glu1104, sont conservés pour les trois conformations. Les types d'interaction apolaires sont aussi bien conservés, mais présentent quelques variations telles que l'ajustement de la chaîne latérale de Cys1083 et les réarrangements continus de la Phe1081, la Leu1150 et la Phe1155. L'état intermédiaire a révélé une nou-
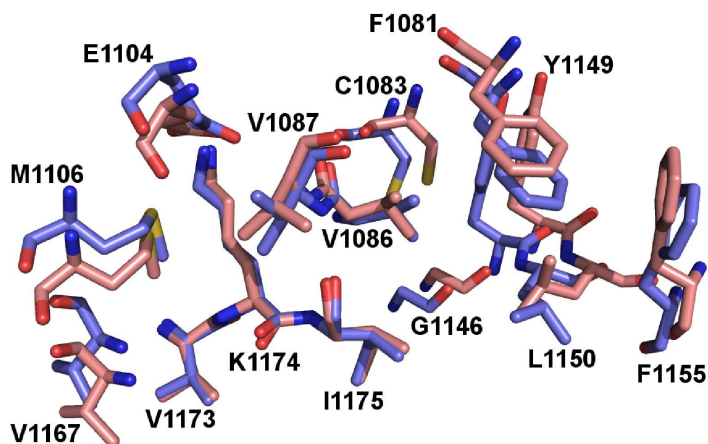
FIG. 6.7: *Une comparaison structurale de l'interface H1-H2 à différents temps de la simulation (0 and 3 ns) pour la conformation productive du DK de l'Insr. Les résidus sont représentés en bâtonnets, étiquetés et colorés selon le type d'atome : L'azote en bleu, l'oxygène en rouge, le soufre en jaune et le carbone en bleu clair pour la conformation à 0 ns et en rouge clair pour celle à 3 ns. Les structures dans l'image sont superposées sur les résidus centraux, la Val1173, la Lys1174 et l'Ile1175.*
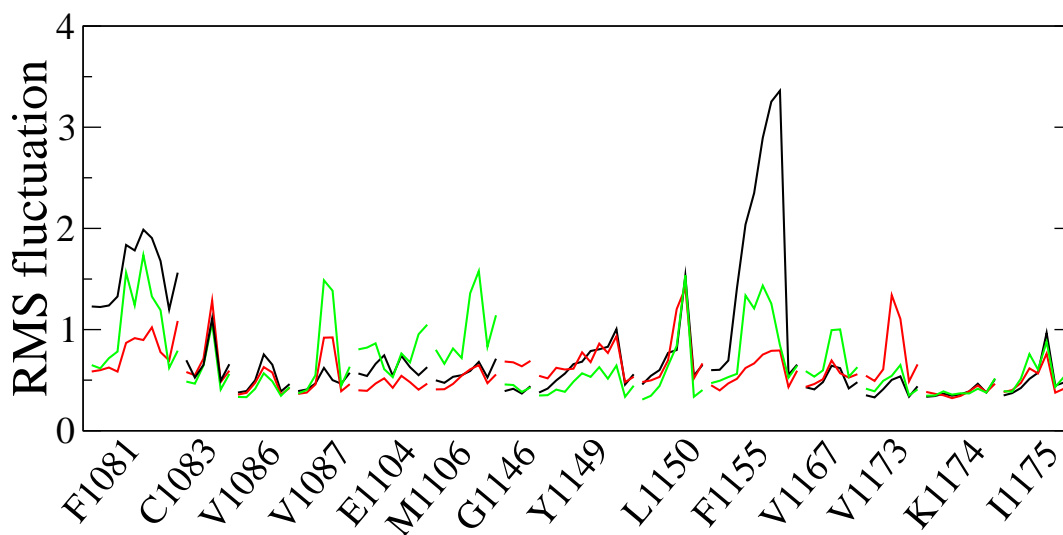

FIG. 6.8: *Les fluctuations RMS des résidus à l'interface entre H1 et H2 des états de l'Insr. Les types de fluctuation en Å sont montrées pour les atomes des résidus et colorées selon un code de couleur : noir pour l'état non-productif, rouge pour le productif et vert pour l'intermédiaire. Les valeurs RMSF aux extrémités du profil pour chaque résidu sont ceux des atomes de la chaîne principale, tandis que les valeurs intermédiaires appartiennent aux chaînes latérales.*

velle organisation des chaînes latérales de la Met1106 et la Val1087.

La conformation dense des résidus de l'interface a été relativement stable pendant la courte durée de simulation. Les variations mineures des liasons hydrophobes pourraient aussi être responsables de mouvements plus grands dans le reste du DK, sur l'échelle de la dynamique naturelle. La rotation de la Cys1083 pourraît être spécifique de la sousfamille des Insr, tandis qu' à la même position, une histidine est hautement conservée. Les contacts hydrophobes de la Met1106 pourraient aussi jouer un rôle dans le comportement dynamique différent comparé à d'autres TKs, car ce résidu est souvent remplacé par des cystéines ou des alanines dont les chaînes latérales sont plus courtes.

Récemment, les structures crystallines du DK de Fgfr2 ont été publiées pour la protéine "wild-type" ainsi que pour quelques protéines mutées [12]. Les résultats présentés pour des mutations du Fgfr2, à la position du Glu1104 et de la Lys1174 selon la numérotation de l'Insr, consolident le rôle critique joué par l'une des deux interactions polaires de l'interface.

Il est possible d'imaginer que les interactions des résidus de l'interface changent pendant que la protéine passe d'un état à l'autre. De tels changements sont partiellement gouvernés par l'hélice C, par le biais de ses interactions avec des résidus de l'interface, notamment Phe1081 et Phe1155 (numérotation selon l'Insr), et son influence sur la conformation de la boucle C-terminale.

Un des problèmes majeurs de la conception *in-silico* des médicaments basée sur la structure est d'avoir une structure protéique comme cible potentielle. Souvent, il n'existe pas de structure tri dimensionnelle adéquate et la modélisation par homologie peut être utilisée pour prédire la conformation voulue d'une protéine. De manière générale, cette méthode identifie les régions de structure conservées (RSCs) d'une structure à transposer dans une séquence cible. L'isolation des RSCs doit être aussi précise que possible pour obtenir un modèle en 3D stable et pertinent. Hormis l'évaluation des structures de l'Insr comme points de départ potentiels pour la modélisation par homologie générale des tyrosine kinases, les résidus de l'interface des noyaux hydrophobes des lobes sont suggérés comme étant des RSCs.

Les découvertes précédentes sur les complexes DK-ligand se sont principalement focalisées sur la définition de 38 résidus définissant le site de liaison de l'ATP [13]. La variabilité des acides aminés au sein les protéines kinases a été étudiée. Des efforts considérables ont été consentis pour tendre à une liaison sélective du ligand. Dans ce cas, la fonction des résidus du noyau de liaison s'avère cruciale à la stabilisation de la conformation du DK. Le centre de liaison des tyrosine kinases intéragit fréquemment avec la phenylalanine du motif DFG dans la forme apo du KD, tandis que ce motif est constamment en contact avec ce centre dans la forme liée. Par conséquent, le noyau de liaison est vu comme le pharmacophore minimal pour les inhibiteurs de la TK compétitifs avec l'ATP.

Plusieurs études sur des structures du DK ont démontré comment l'interaction et la conformation de trois éléments du DK, notamment l'hélice C, la boucle riche en glycines et la boucle A, commandent l'arrangement structural de tyrosine kinases spécifiques dans certains états [14–16]. Dans ce travail, une autre approche à été retenue, afin de comprendre un nombre important de conformations de différentes TKs. L'hypothèse émise est qu'il devrait y avoir des bouts d'acides aminés dans la structure protéique primaire du DK dont les variations entraînent différentes topologies tridimensionnelles, conformément à ce que l'on observe dans les structures expérimentales. Les résidus de l'interface H1-

H2 ont été révélés comme une région potentiellement impliquée dans la stabilisation des conformations atteignables et dans la dynamique du domaine catalytique. Néanmoins, elle n'est pas l'unique région présentant de telles caractéristiques, mais elle est certainement importante et jusqu'à maintenant elle a peu été explorée.

La description moléculaire de l'activation du DK à déjà été essayée auparavant [17]. Dans ces études, de manière générale, les changements conformationnels sont forcés par l'application de forces externes ou de potentiels au système. La nouveauté des essais présentés est la simulation d'un système auto-inhibiteur perturbé. La réaction du DK de Flt3 jusqu'au relâchement du mécanisme auto-inhibiteur à été suivie par dynamique moléculaire. L'interêt était de souligner quels éléments de la protéine subissent un mouvement tandis que le système va d'un état thermodynamique à un autre. L'approche a été couronnée de succès et a fourni une dynamique de l'hélice C jouant un rôle majeur dans le changement conformationnel du lobe N-terminal, lequel est censé être la premiére étape de la cascade d'activation.

Bibliographie

[1] W L DeLano. The PyMOL molecular graphics system. 2003 at http//www.pymol.org .

[2] S K Hanks and T Hunter. Protein kinases 6. The eukaryotic protein kinase superfamily : kinase (catalytic) domain structure and classification. *FASEB J*, 9(8) : 576-596, May 1995.

[3] Y Segawa, H Suga, N Iwabe, C Oneyama, T Akagi, T Miyata and M Okada. Functional development of Src tyrosine kinases during evolution from a unicellular ancestor to multicellular animals. *Proc Natl Acad Sci U S A*, 103(32) : 12021-12026, Aug 2006.

[4] G Manning, G D Plowman, T Hunter and S Sudarsanam. Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci*, 27(10) : 514-520, Oct 2002.

[5] T Pawson. Protein modules and signalling networks. *Nature*, 373(6515) : 573-580, Feb 1995.

[6] T Hunter. The Croonian Lecture 1997. The phosphorylation of proteins on tyrosine : its role in cell growth and disease. *Philos Trans R Soc Lond B Biol Sci*, 353(1368) : 583-605, Apr 1998.

[7] G Manning, D B Whyte, R Martinez, T Hunter and S Sudarsanam. The protein kinase complement of the human genome. *Science*, 298(5600) : 1912-1934, Dec 2002.

[8] P Blume-Jensen and T Hunter. Oncogenic kinase signalling. *Nature*, 411(6835) : 355-365, May 2001.

[9] B Nagar, W G Bornmann, P Pellicena, T Schindler, D R Veach, W T Miller, B Clarkson and J Kuriyan. Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res*, 62(15) : 4236-4243, Aug 2002.

[10] R Capdeville, E Buchdunger, J Zimmermann and A Matter. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat Rev Drug Discov*, 1(7) : 493-502, Jul 2002.

[11] C D Mol, D R Dougan, T R Schneider, R J Skene, M L Kraus, D N Scheibe, G P Snell, H Zou, B C Sang, K P Wilson. Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *J Biol Chem*, 279(30) : 31655-31663, Jul 2004.

[12] H Chen, J Ma, W Li, A V Eliseenkova, C Xu, T A Neubert, W T Miller and M A Mohammadi. A molecular brake in the kinase hinge region regulates the activity of receptor tyrosine kinases. *Mol Cell*, 27(5) : 717-730, Sep 2007.

[13] A Vulpetti and R Bosotti. Sequence and structural analysis of kinase ATP pocket residues. *Farmaco*, 59(10) : 759-765, Oct 2004.

[14] D A Johnson, P Akamine, E Radzio-Andzelm, M Madhusudan and S S Taylor. Dynamics of cAMP-dependent protein kinase. *Chem Rev*, 101(8) : 2243-2270, Aug 2001.

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

[16] S S Taylor, J Yang, J Wu, N M Haste, E Radzio-Andzelm and G Anand. PKA : a portrait of protein kinase dynamics. *Biochim Biophys Acta*, 1697(1-2) : 259-269, Mar 2004.

[17] E Ozkirimli and C B Post. Src kinase activation : A switched electrostatic network. *Protein Sci*, 15(5) : 1051-1062, May 2006.

Curriculum vitae

NAME	Loris Moretti
CITIZENSHIP	Italian
BIRTH	December 9th 1975, Monselice/Padova (Italy)
EMAIL	laurenziano@inwind.it

Education

2004-present
Department of Pharmaceutical Sciences, University of Geneva (CH)
Ph.D. studies at University of Geneva (CH), ETH Zurich (CH) and
University of Padova (IT)
Co-supervised by Prof. Leonardo Scapozza and Prof. Stefano Moro

2003
Department of Medicinal Chemistry, University of Padova (IT)
Visiting researcher at Stefano Moro's group at MMS-Lab

2002
Department of Biochemistry and Pharmacy, Åbo Akademi University (FIN)
Post-lauream fellowship at the Mark Johnson's group

2002
Department of Medicinal Chemistry, University of Padova (IT)
M.S. degree in Pharmaceutical Chemistry and Technology at University of Padova
Co-supervised by Prof. Stefano Moro and Prof. Mark Johnson

2001
Department of Biochemistry and Pharmacy, Åbo Akademi University (FIN)
Erasmus scholarship at the Mark Johnson's group

1994
High School degree, Monselice (IT)

Research fields & skills

Theories and applications
Molecular modeling, bioinformatics, drug design, data modeling and
molecular dynamics simulation

Tools
Bodil, Gromacs, Amber, Sybyl, Moe, Chimera, Corina, ClustaW, Vida2, filter, Gold,
Dock, Flexx, etc...
Linux operating system, as administrator
Shell scripting

Languages

English, Finnish (basic level), French, German (basic level), Italian (mother tongue)

Invited lecture

- Second joint Italian-Swiss meeting on medicinal chemistry, Modena, (IT); "Tyrosine
  kinase drug discovery: what can be learned from solved crystal structures?"

Posters

2007

- Frontiers in medicinal chemistry. Berlin, (D); "An Imatinib chloro-derivative shows
  long-lasting effects" Symphosium of union of the Swiss societies for experimental
  biology, USEGEB. Geneva, (CH); "MD investigation about the implication of ITD
  of FLT3 in oncogenesis"

2006

- Protein kinase inhibitors. Paris, (F); "The ligand core of TK inhibitors and its appli-
  cation on anaplastic lymphoma kinase"

- Helsinki drug design. HDR. Helsinki, (FIN); "Beyond the ligand core: possibilities of
  selectivity for the tyrosine kinase inhibitors"

- Symphosium of union of the Swiss societies for experimental biology, USEGEB. Geneva, (CH); "Exploiting experimental data for tyrosine kinase drug discovery through molecular modeling"

    2004

- EuroQSAR, The 15th European Symposium on Quantitative Structure-Activity Relationships & Molecular Modeling, Istanbul, (TR); "Exploring SKI-606 Ligand Conformations Through Molecular Docking Study on Abl TK"

Publications

- Puttini M., Redaelli S., **Moretti L.**, Brussolo S., Gunby R., Mologni L., Marchesi E, Cleris L., Donella-Deana A., Drueckes P., Sala E, Lucchini V., Kubbutat M., Formelli F., Zambon A., Scapozza L. and Gambacorti-Passerini C. Characterization of CGP-584, a new Abl-inhibitor with long-lasting effects. Submitted

- **Moretti L.**, Tchernin L. and Scapozza L. Tyrosine kinase drug discovery: what can be learned from solved crystal structures? ARKIVOC. (2006) 8: 38-49.

- Grazioso G., **Moretti L.**, Scapozza L., De Amici M., De Micheli C. Development of a three-dimensional model for the N-methyl-D-aspartate NR2A subunit. J Med Chem. (2005) 48(17):5489-94.

- **Moretti L.**, and Scapozza L. (2004) Exploring SKI-606 Ligand Conformations Through Molecular Docking Study on Abl Tyrosine Kinase. Proceedings book of the EuroQSAR 2004.

- **Moretti L.**, Pentikäinen O.T., Settimo L. and Johnson M.S. (2004) Model structures of the N-methyl-D-aspartate receptor subunit NR1 explain the molecular recognition of agonist and antagonist ligands. Journal of Structural Biology 145: 205-215.

Acknowledgments

me to understand the computers' point of view.

If I could arrived at this point it is thank to all the people I acknowledged above and to many others who I have not metioned but they know what they meant to me . . .

Thanks . . .