
(SSD ING-IND/24 – Principi di Ingegneria Chimica)

DOTTORANDO
Giovanni Maria Pavan

DIRETTORE DELLA SCUOLA
Chiar.mo Prof. Maurizio Fermeglia (Università di Trieste)

RELATORE
Chiar.mo Prof. Sabrina Pricl (Università di Trieste)

ANNO ACCADEMICO 2008/2009
Ai miei genitori,
ai miei fratelli, e ai sogni... e a chi mi aiuta a raggiungerli...
Abstract (English)

The notable progresses achieved in the medical biotechnology research allowed to identify the common genetic origin of many different pathologies. Genes that, for unknown reasons, mutate during the lifetime of the patient synthesize proteins and receptors that lose the control by grow factors. These aberrant proteins are consequently always active and give rise to a series of cascade signals that result in an uncontrolled proliferation of tumor cells. In this framework, the use of small inhibitor molecules to deactivate these proteins, and consequently to block the diseases, constitutes a revolutionary concept that is the base of the “target therapy”. However, during the treatment, patients develop almost unavoidably resistance toward these kind of drugs that is caused by mutations that interest the amino acidic chains of determined tyrosine kinases (i.e. KIT, PDGFRα, etc.).

In this thesis we used molecular modelling techniques to gain an insight in the binding between proteins and inhibitors. In particular, we were able to obtain unique information about the interactions that stabilize the complexation, and consequently about the destabilizing effect induced by mutations in the protein and receptor structures. Our data show that different drugs penetrate differently inside the binding site of kinases during the inhibition. Consequently, the efficiency of each inhibitor molecule is strongly dependent on the situation (mutations) presented by each patient. It becomes thus necessary to create a benchmark of responses related to all the most common mutations treated with different drugs. In this direction, molecular simulation can really support clinical scientists in the comprehension of drug resistance phenomena – a virtual microscope, able to provide unique details to cancer research.
Abstract (Italiano)

I notevoli progressi compiuti dalla ricerca nelle biotecnologie mediche ha permesso di identificare la comune origine genetica di diverse patologie. Geni che, per cause ancora sconosciute, mutano durante la vita del paziente, sintetizzano proteine e recettori che perdono il controllo da parte dei fattori di crescita. Queste proteine aberranti sono di conseguenza sempre attive e generano una serie di segnali “a cascata” che risultano in una proliferazione incontrollata delle cellule tumorali. In questo ambito, l’uso di piccole molecole inibitrici per deattivare queste proteine, bloccando di conseguenza la progressione delle malattie, costituisce il concetto rivoluzionario che sta alla base della terapia a “bersaglio molecolare” (target therapy). Durante il trattamento tuttavia i pazienti sviluppano quasi inevitabilmente una sorta di resistenza nei confronti dei questo tipo di farmaci che è causato dalla comparsa di mutazioni che interessano la catena aminoacidica di determinate chinasi (KIT, PDGFRα, ecc.).

In questa tesi sono state utilizzate tecniche di modellistica molecolare per ottenere un’analisi dettagliata del binding tra proteine e inibitori. In particolare, è stato possibile ottenere importanti informazioni sulle interazioni che stabilizzano la complessazione, e di conseguenza sull’effetto destabilizzante generato da mutazioni che compaiono sulla struttura di proteine e recettori. I nostri risultati mostrano che farmaci diversi penetrano differentemente all’interno del sito di binding delle chinasi durante l’inibizione. L’efficienza di ciascun inibitore risulta quindi essere fortemente dipendente dalla situazione (mutazioni) del singolo paziente. Diviene dunque necessaria la creazione di un benchmark contenente i responsi di tutte le più comuni mutazioni trattate con diversi farmaci. In questa direzione, la simulazione molecolare può fornire un supporto reale alla clinica nella comprensione dei fenomeni di resistenza ai farmaci: un microscopio virtuale, capace di fornire dettagli unici alla ricerca sul cancro.
# Table of Contents

1. **Introduction** .................................................................................................................. 5  
   1.1. Introduction .................................................................................................................. 6  
   1.2. The diseases and their origin ...................................................................................... 7  
      1.2.1. The gastrointestinal stromal tumor (GIST) .............................................................. 7  
      1.2.2. The melanoma ........................................................................................................ 9  
      1.2.3. HES and CEL ......................................................................................................... 10  
   1.3. Tyrosine-kinases .......................................................................................................... 12  
      1.3.1. The structure of the tyrosine kinase domain ........................................................... 12  
      1.3.2. The signal and its inhibition .................................................................................... 13  
   1.4. The inhibitor drugs ....................................................................................................... 17  
      1.4.1. Imatinib .................................................................................................................. 18  
      1.4.2. Dasatinib ............................................................................................................... 19  
      1.4.3. Sorafenib, nilotinib and sunitinib .......................................................................... 20  
   1.5. Mutations and drug resistance .................................................................................... 21  
      1.6. The role of molecular simulation .............................................................................. 23  

2. **Materials & Methods** .................................................................................................... 25  
   2.1. Introduction: the molecular simulation ....................................................................... 26  
   2.2. The atomistic scale of simulation .............................................................................. 26  
   2.3. Software, force field and parametrization .................................................................. 31  
   2.4. Creation of “wild type” complexes .......................................................................... 34  
   2.5. Creation of mutated complexes .................................................................................. 47  
      2.5.1. Point mutations ..................................................................................................... 48  
      2.5.2. Aminoacidic deletions .......................................................................................... 48  
   2.6. The simulation procedure .......................................................................................... 50  
   2.7. Affinity energy analysis (MM-PBSA) ........................................................................ 55  
   2.8. Per-residue decomposition analysis .......................................................................... 59
3. Results & Discussion

3.1. Preface

3.2. The T670I mutation in KIT

3.2.1. KIT T670I, Bcr-Abl T315I and PDGFRα T674I

3.2.2. T670X: the results

3.3. Study of two new mutations in PDGFRα

3.3.1. PDGFRα model, point mutation and aminoacidic deletion

3.3.2. Results and discussion

3.4. Synergetic effect between primary and secondary mutations in KIT

3.4.1. Primary and secondary mutations

3.4.2. Results and discussion

3.5. Characterization of L576P mutation in the KIT-imatinib complex

3.5.1. Mutations in the juxtamembrane domain

3.5.2. History of the cases studied

3.5.3. Results

3.6. Inhibition of KIT L576P: imatinib vs. dasatinib

3.6.1. Introduction

3.6.2. A different functionality: the results

3.6.3. Inside the binding site

3.7. KIT L576P: the functionality of different inhibitors

3.7.1. Two different penetrations

3.7.2. Outcomes

3.8. KIT mutations: benchmark of affinity toward imatinib

3.8.1. Early results

4. Conclusions

References
Chapter 1

Introduction

This Chapter reports background, motivations and objectives of the present research. In this section, we will describe the systems that will be investigated in this thesis and we will introduce the base principles of the target therapy and of the molecular treatment of the pathologies of interest.
1.1. Introduction

The notable progresses achieved during the last century in the genetic and medical biotechnology research allowed, since the 1960s, to identify the chromosome anomaly that is responsible of the chronic myeloid leukemia (CML). Due to the numerous studies and efforts on this topic, it is presently well known that, in most of cases, this pathology is caused by the Philadelphia chromosome that express an oncogene (BCR-ABL) that is not present in the healthy patients. This gene syntethize an aberrant protein, the Abelson protein (Bcr-Abl) whose tyrosine-kinase domain is costitutively always active and phosphorylated. The consequence is a series of “cascade signals” that results in the uncontrolled reproduction of the tumoral cells. The traditional therapies demonstrated to be poorly effective in eradicating definitely such pathologies. Since 2002 a new drug, the imatinib (or Gleevec®), was introduced into the market, showing exceptional efficiency in binding to the Bcr-Abl tyrosine-kinase domain, inhibiting its activity.

Due to the encouraging results achieved for LMC, recently scientists started to test the efficacy of this drug also toward different diseases that have a genetic origin very close to the one of CML, and that are the subjects of this thesis – i.e. the gastrointestinal stromal tumors (GIST), the hypereosinophilic idiopatic syndrome (HES), the chronic eosinophilic leukemia (CEL) as well as the melanoma. The characteristic of these pathologies have in common with CML is that they are driven by the presence in the organism of different aberrant tyrosine-kinases – the KIT receptor and the PDGFRα protein, that, similarly to Bcr-Abl (in CML), express an uncontrolled tyrosine-kinase activity. That is why these pathologies are ideal candidates for the “target therapy” treatment.

During the therapy with imatinib, however, a relapse of these tumors was evidenced, indicating a loss of efficiency of the drug action toward all the kinases. Clinical analyses on the patients that expressed resistance to imatinib evidenced the presence of mutations (point mutations, amino acid deletions and duplications) in the amino acid chains of KIT, PDGFRα as well as of Bcr-Abl. Further studies evidenced the presence of point mutations in the BCR-ABL, c-kit and FIP1L1-PDGFRα oncogenes. The related proteins presented consequently a single amino acid substitution respect to the “wild type” case that was, however, able to affect dramatically the binding with the drug.
In the works presented in this thesis molecular modelling techniques are used to analyze the interactions between proteins characteristic of different pathologies and various inhibitor drugs. By comparing the affinity of binding for the drug of the “wild type” protein with the one expressed by the mutated structures it is possible to evaluate the destabilizing effect induced by the substitution of a single residue that is related to this loss of functionality effect.

1.2. The diseases and their origin

In this section we give a description of the main characteristics of the pathologies related to the proteins analyzed in this thesis – gastrointestinal stromal tumors (GISTs), melanoma, hypereosinophilic idiopathic syndrome (HES) and chronic eosinophilic leukemia (CEL). As it will be highlighted in the further paragraphs, these diseases present common aspects that make them ideal candidates for the target therapy.

1.2.1. The gastrointestinal stromal tumor (GIST)

A gastrointestinal stromal tumor (GIST) is one of the most common mesenchymal tumors of the gastrointestinal tract (1-3% of all gastrointestinal malignancies). This term identifies an extremely heterogeneous number of tumors that interest the gastrointestinal tract. They are typically defined as tumors whose behavior is driven by mutations in the c-kit gene or PDGFRA gene.\(^1\)

GISTs are tumors that interest the connective tissue – i.e. sarcomas. Unlike most gastrointestinal tumors, they are non-epithelial diseases. Among GISTs, the 70% occur in the stomach, 20% in the small intestine and less than 10% in the esophagus. Small tumors are generally benign, but large tumors can interest the liver and peritoneal cavity. They rarely occur in other abdominal organs. GISTs arise from interstitial cells of Cajal (ICC),\(^2\) that are normally part of the autonomic nervous system of the intestine and control motility. Most (50-80%) GISTs arise because of a mutation in a gene called c-kit. This gene encodes a transmembrane receptor for a growth factor termed scf (stem cell factor). The KIT product/CD117 is expressed on ICCs and a large number of other cells, mainly
bone marrow cells, mast cells, melanocytes and several others. The KIT molecule comprises a long extracellular domain, a transmembrane segment, and an intracellular part. Mutations generally occur in the DNA encoding the intracellular part (exon 11), which acts as a tyrosine kinase to activate other enzymes. Mutations make KIT function independent of activation by scf, leading to a high cell division rate and possibly genomic instability. It is likely that additional mutations are "required" for a cell with a KIT mutation to develop into a GIST, but the c-kit mutation is probably the first step of this process.

Figure 1.1. Structure the KIT receptor (Protein Data Bank code: 1T46). Within the structure, α-helix are colored in red and β-sheets in cyan.
Until now, GISTs were notorious for being resistant to chemotherapy, with a success rate of <5%. Recently, the KIT tyrosine kinase inhibitor imatinib (Glivec/Gleevec), a drug initially marketed for CML, was found to be useful in treating GISTs, leading to a 40-70% response rate in metastatic or inoperable cases. Moreover, patients who develop resistance to imatinib was found to possibly respond to the different tyrosine kinase inhibitor – i.e. sunitinib (Sutent).³

1.2.2. The melanoma

Melanoma is a malignant tumor of melanocytes which are found predominantly in skin but also in the bowel and the eye. It is one of the less common types of skin cancer but causes the majority (75%) of skin cancer related deaths.⁴ Melanocytes are normally present in skin, being responsible for the production of the dark pigment melanin.⁵ Despite many years of intensive laboratory and clinical research, the greatest chance of cure is in the early surgical resection of thin tumors.

Familial melanoma is genetically heterogeneous,⁶ and a common origin for familial melanoma have been identified on the chromosome arms 1p, 9p and 12q. Multiple genetic events have been related to the pathogenesis of melanoma.⁷ Today, melanomas are diagnosed only after they become visible on the skin. Recent genetic advances however will allow to support doctors in the identification of people with high-risk genotypes and to determine the cases that have the greatest chance of becoming malignant cancerous. A number of rare mutations, which often run in families, are known to greatly increase the susceptibility to melanoma. One class of mutations affects the gene CDKN2A. An alternative reading frame mutation in this gene leads to the destabilization of p53, a transcription factor involved in apoptosis and in fifty percent of human cancers. Another mutation in the same gene results in a non-functional inhibitor of CDK4, a [cyclin-dependent kinase] that promotes cell division. Scattered throughout the genome, these mutations reduce a cell’s ability to repair DNA. Both CDKN2A and XP mutations are highly penetrant.

Patients with metastatic melanoma have a median survival of 6 to 8 months.⁸ Unfortunately multiple clinical trials with chemotherapy, immunotherapy, and
biochemotherapy have failed to significantly improve survival. Protein kinase inhibitors are beneficial in diseases with highly prevalent oncogenic events (e.g., CML, GIST), and within selected subpopulations (e.g., HER2-amplified breast cancer). More than 50% of melanomas arising from areas without chronic sun damage harbor activating mutations of BRAF. Interestingly, amplification of chromosomal region 4q12 was seen frequently and selectively in AL, CSD, and mucosal melanomas. Interrogation of candidate genes in this region led to the discovery of frequent mutations and or amplifications of the KIT tyrosine kinase receptor gene in these subtypes. The identification of KIT mutations in melanoma has direct therapeutic implications. Activating KIT mutations are present in about 85% to 90% of gastrointestinal stromal tumors (GIST). Treatment with the KIT inhibitor imatinib significantly improved survival in GIST patients, and it is now the standard of care for this disease. For this reason, this drug can be successfully adopted also for the treatment for malignant melanoma showing a KIT receptor deregulation that is caused by the presence of amino acid mutations.

1.2.3. HES and CEL

The idiopathic hypereosinophilic syndrome (HES) and the chronic eosinophilic leukemia (CEL) are two pathologies that interest the eosinophil cells. In details, HES is a disease characterized by a persistently elevated eosinophil count (≥ 1500 eosinophils/mm³) in the blood for at least six months without any recognizable cause, with involvement of either the heart, nervous system, or bone marrow. There are some associations with chronic eosinophilic leukemia (CEL) as it shows similar characteristics and genetic defects.

Similarly, the chronic eosinophilic leukemia (CEL) is a highly rare disease in which too many eosinophils are found in the bone marrow, blood, and other tissues. CEL may stay the same for many years, or it may progress quickly to acute leukemia. It is generally caused by overactivation of the oncogene PDGFRA through a chromosome translocation event. Similarly to the pathologies described previously, also CEL is extremely manageable with the use of Gleevec, which suppresses the oncogenic effects of PDGFRA.
Figure 1.2. Structure the PDGFRα (its structure is not present in the Protein Data Bank and was created by homology). Within the structure, α-helix are colored in red and β-sheets in cyan.
1.3. Tyrosine-kinases

A tyrosine kinase is an enzyme that can transfer a phosphate group from ATP to a tyrosine residue in a protein. Tyrosine kinases are a subgroup of the larger class of protein kinases. Phosphorylation of protein kinases is an important mechanism in signal transduction for regulation of enzyme activity. As said, the kinase structures of the proteins that are characteristic of the pathologies described previously present many analogies. This particular family of proteins represent the ideal candidates for the therapy with inhibitor drugs – the main purpose is to stop the signal that result by an uncontrolled phosphorylation.

1.3.1. The structure of the tyrosine-kinase domain

Approximately 2000 kinases are known, and more than 90 Protein Tyrosine Kinases (PTKs) have been found in the human genome. They are divided into two classes, receptor and non-receptor PTKs. Presently, 58 receptor tyrosine kinases (RTKs) are known, grouped into 20 subfamilies. They play a crucial roles in many cellular activities – i.e. in the growth, differentiation, metabolism, adhesion, motility, death. RTKs are composed of an extracellular domain, which is able to bind a specific ligand, a transmembrane domain, and an intracellular catalytic domain, which is able to bind and phosphorylate selected substrates. The binding of a ligand to the extracellular region causes a series of structural rearrangements in the RTK that results in enzymatic activation. In particular, movement of some parts of the kinase domain gives free access to adenosine triphosphate (ATP) and the substrate to the active site. This results in a cascade of events through phosphorylation that transmit ("transduce") the extracellular signal to the nucleus, causing changes in gene expression. Many RTKs are involved in oncogenesis, either by gene mutation, or chromosome translocation, or simply by over-expression. In any case, the result is a hyper-active kinase, that confers an aberrant, ligand-independent, non-regulated growth stimulus to the cancer cells.

Tyrosine kinases are particularly important today due to their implications in the treatment of cancer. A mutation that causes certain tyrosine kinases to be constitutively active has been associated with several cancers. Presently, there are various drugs (i.e.
imatinib, dasatinib, ecc.) able to bind to the binding site of these tyrosine kinases, inhibiting its activity.\textsuperscript{18}

Figure 1.3. Structure of a tyrosine-protein kinase both in open (a: active, Lck) and in closed (b: inactive, Src) configuration.

1.3.2. The signal and its inhibition

Protein kinases are a group of enzymes that possess a catalytic subunit which transfers the gamma phosphate from nucleotide triphosphates (often ATP) to one or more amino acid residues in a protein substrate side chain, resulting in a conformational change affecting protein function. The enzymes can be divided into two classes: serine/threonine specific and tyrosine specific.\textsuperscript{19}

Phosphorylation is the addition of a phosphate (PO4) group to a protein or other organic molecule. This process activates or deactivates many protein enzymes, causing or preventing the mechanisms of diseases such as cancer or different pathologies. Protein phosphorylation in particular plays a significant role in a wide range of cellular processes.
Many enzymes and receptors are switched "on" or "off" by phosphorylation and dephosphorylation. Reversible phosphorylation results in a conformational change in the structure in many enzymes and receptors, causing them to become activated or deactivated. Phosphorylation usually occurs on serine, threonine, and tyrosine residues in eukaryotic proteins. In addition, phosphorylation occurs on the basic amino acid residues histidine or arginine or lysine in these kind of proteins.\textsuperscript{20,21}

![Figure 1.4. The phosphorylation process and the consequent signal of cell replication (green arrow).](image)

The addition of a phosphate (PO\(_4\)) molecule to a polar R group of an amino acid residue can turn a hydrophobic portion of a protein into a polar and extremely hydrophilic portion of molecule. In this way it can introduce a conformational change in the structure
of the protein via interaction with other hydrophobic and hydrophilic residues in the protein. One such example of the regulatory role that phosphorylation plays is the p53 tumor suppressor protein. The p53 protein is heavily regulated\textsuperscript{22} and contains more than 18 different phosphorylation sites. Activation of p53 can lead to apoptotic cell death.\textsuperscript{23} This activity occurs only in situations of cell damage or when the physiology is disturbed in normal healthy individuals.

Figure 1.5. The p53 tumor suppressor protein in complex with DNA.
Upon the deactivating signal, this protein becomes dephosphorylated again and stops working. This is the mechanism called “signal transduction” that is common also for the tyrosine kinases.

When, as said, for some reasons, the genes that synthesize these proteins or receptor is mutated, it generates proteins and receptors (i.e. KIT) that are constitutively always active and phosphorylated – they lose the control by grow factors. The result is an uncontrolled proliferation of the tumoral cells that must be stopped in order to block the disease. Inhibitor drugs are a breakthrough in this direction since they are able to block the origin of the tumoral growth – the tyrosine kinase signal.

Figure 1.6. Mechanism of the tyrosine-kinase inhibition by an inhibitor drug (i.e. imatinib, dasatinib, ecc).

Figure 1.6 represents the inhibition mechanism by a drug like imatinib. The inhibitor molecule penetrates within the binding site of the protein (or receptor) occupying the position that usually is reserved to ATP. It definitely prevents the penetration by ATP, stopping the activity of the protein.
In this thesis different drugs will be considered. In fact, different molecules have different structures and thus different penetrations inside the binding pocket of the protein. The works presented here will highlight how important this functionality is, and how different drugs can be effective or useless for the therapy, depending on the situation presented by each patient.

1.4. The inhibitor drugs

In this section we will provide a description of the inhibitor drugs studied in this thesis. As anticipated in the previous paragraph, they can show different functionality – and thus efficacy – depending on the chemical structure.

Figure 1.7. The Bcr-Abl protein in complex with the imatinib drug (red).
1.4.1. Imatinib

Imatinib is a drug used to treat certain types of cancer. It is currently produced by Novartis as Gleevec (USA) or Glivec (Europe/Australia) as its mesylate salt, imatinib mesilate (INN) (formerly called STI-571). It is used in treating chronic myelogenous leukemia (CML), gastrointestinal stromal tumors (GISTs) and other cancers.

It is the first member of this new class of agents that act by specifically inhibiting a certain enzyme that is characteristic of a particular cancer cell, rather than non-specifically inhibiting and killing all rapidly dividing cells.

In CML, the enzyme tyrosine kinase is stuck in the "on" position. Imatinib binds to the site of tyrosine kinase activity, and prevents its activity.

Imatinib is used in chronic myelogenous leukemia (CML), gastrointestinal stromal tumors (GISTs) and a number of other malignancies. One study demonstrated that Imatinib mesylate was effective in patients with systemic mastocytosis, including those who had the D816V mutation in KIT. Experience has shown, however, that imatinib is much less effective in patients with this mutation, and patients with the mutation comprise nearly 90% of cases of mastocytosis. Early clinical trials also show its potential for treatment of hypereosinophilic syndrome and dermatofibrosarcoma protuberans.

Imatinib may also have a role in the treatment of pulmonary hypertension. It has been shown to reduce both the smooth muscle hypertrophy and hyperplasia of the pulmonary vasculature in a variety of disease processes, including portopulmonary hypertension.

In laboratory settings, imatinib is being used as an experimental agent to suppress platelet-derived growth factor (PDGF) by inhibiting its receptor (PDGFRα). One of its effects is delaying atherosclerosis in mice with diabetes.

Figure 1.8. The imatinib structure.
Imatinib is a 2-phenylaminopyrimidine derivative that functions as a specific inhibitor of a number of tyrosine kinase enzymes. It occupies the TK active site, leading to a decrease in activity.

There is a large number of tyrosine kinase enzymes in the body, including the insulin receptor. Imatinib is specific for the binding domain in Abl (the Abelson proto-oncogene), KIT and PDGFRα (platelet-derived growth factor receptor). The active sites of tyrosine kinases each have a binding site for ATP. The enzymatic activity catalyzed by a tyrosine kinase is the transfer of the terminal phosphate from ATP to tyrosine residues on its substrates, a process known as protein tyrosine phosphorylation. Imatinib works by binding to the ATP binding site of Bcr-Abl and inhibiting the enzyme activity of the protein competitively.27

1.4.2. Dasatinib

Dasatinib, also known as BMS-354825, is a cancer drug produced by Bristol-Myers Squibb and sold under the trade name Sprycel. It is an oral dual BCR/ABL and Src family tyrosine kinases inhibitor approved for use in patients with (CML) after imatinib treatment and Philadelphia chromosome-positive acute lymphoblastic leukemia (Ph+ ALL).

Figure 1.9. The dasatinib structure.
Recently (in one of the studies reported in this thesis) it was found to be efficient also in the inhibition of the KIT receptor, in the treatment of metastatic melanoma.\textsuperscript{28}

### 1.4.3. Sorafenib, nilotinib and sunitinib

In this paragraph three different drugs of more recent introduction in the treatment of these tumors will be treated. They are sorafenib, nilotinib and sunitinib and, as will be described in details in the Result & Discussion chapter, present some common functionalities with the more widely used imatinib and dasatinib.

Sorafenib (marketed as Nexavar by Bayer), is a drug approved for the treatment of primary kidney cancer (advanced renal cell carcinoma) and advanced primary liver cancer (hepatocellular carcinoma). It is a small molecular inhibitor of several tyrosine protein kinases\textsuperscript{29} that are active in many of the molecular pathways that cause cells to become cancerous. These pathways include Raf kinase, PDGF (platelet-derived growth factor), VEGF receptor 2 and 3 kinases and KIT, the receptor for Stem cell factor. A growing number of drugs target most of these pathways.

Nilotinib, in the form of the hydrochloride monohydrate salt, is a tyrosine kinase inhibitor of BCR-ABL.\textsuperscript{30} It was approved as Tasigna in the USA and the EU for drug-resistant chronic myelogenous leukemia (CML).\textsuperscript{31} In June 2006, a Phase I clinical trial found nilotinib to have a relatively favorable safety profile and to show activity in cases of CML resistant to treatment with imatinib.\textsuperscript{32} In that study 92\% of patients (already resistant or unresponsive to imatinib) achieved a normal white blood cell counts after five months of treatment.\textsuperscript{33} The drug was demonstrated, however, to potentially create possible heart complications.\textsuperscript{34,35}

As last, sunitinib (marketed as Sutent, and previously known as SU11248) is an oral, small-molecule, multi-targeted receptor tyrosine kinase (RTK) inhibitor that was approved by the FDA for the treatment of renal cell carcinoma (RCC) and imatinib-resistant gastrointestinal stromal tumor (GIST) in 2006. It was the first cancer drug simultaneously approved for two different indications.\textsuperscript{36} Sunitinib has become a standard of care for both of these cancers, and is currently being studied for the treatment of many others.
Sunitinib is able to inhibit VEGFRs, PDGFRα and KIT\(^{37}\) that, when improperly activated by mutations, drives the majority of gastrointestinal stromal cell tumors.\(^{38}\) It has been recommended as a second-line therapy for patients whose tumors develop mutations in KIT that make them resistant to imatinib, or who become intolerant to the drug.\(^{39,40}\) Approximately 20% of patients do not respond to imatinib (early or primary resistance), and among those who do respond initially, 50% develop secondary imatinib resistance and disease progression within 2 years. Prior to sunitinib, patients had no therapeutic option once they became resistant to imatinib.\(^{41}\)

The efficacy of sunitinib is currently being evaluated in a broad range of solid tumors, including breast, lung, and colorectal cancers. Early studies have shown single-agent efficacy in a number of different areas.

### 1.5. Mutations and drug resistance

Even if the use of these molecule to inhibit the functionality of these onco-proteins is a breakthrough in the treatment of many tumors, and the target therapy represents a
promising technique especially in terms of efficiency and side effects, many problems in this direction must still be solved.

As a first, it is worth noting that patients affected by CML, GISTs, HES, CEL and melanoma treated for instance with imatinib, presented after 2-6 months in average a consistent relapse.

From the molecular point of view the relapse is characterized with the reappearance of the aberrant proteins – the inhibitor drug loses efficiency in its action. Substantially, this can happen because of the amplification of the BCR-ABL, c-kit e FIP1L1-PDGFRα genes, or due to their mutation. As said, mutated genes can codify mutated proteins and receptors. These mutations, can modify the thermodynamic and energetic stability of the protein-drug complex. In practice, due to these mutations, proteins and drugs do not fit perfectly anymore. These are the most common cases in the imatinib resistance phenomena. For example, the most known destabilizing mutations in the proteins of interest in this study versus the imatinib drug are the T670I in KIT and the T674I in PDGFRα, that are, as will be described further, not only homologues, but also both homologues of the T315I in Bcr-Abl, well known for its dramatic resistance toward the inhibition with imatinib.

Figure 1.11. Van der Waals surfaces of 315 residues toward the imatinib drug. In the “wild type” protein, the 315 residue is a threonine (blue) and in the mutated case it is an
isoleucine (red). It is evident how the higher steric occupancy of the isoleucine create a conflict with the one of the drug. (Picture taken form: Gorre ME, Sawyers CL. *Curr. Opin. Hematol.* 2002; 9(4): 303-307).

Mutations can be divided into point mutations (substitution of an amino acid with another one), amino acid deletions and duplications. They can either be separated in two categories – primary and secondary mutations (details on the case studied will be discussed in the Results & Discussion section. Secondary mutations (the most common in the case of imatinib resistance) are the ones that appear during the treatment with the drug, creating resistance phenomena and the loss of affinity for the inhibitor molecule.

In this framework, it is really important the different functionality of these drugs. As it will be clarified in the further sections, different penetration of a molecule within the binding site of a protein can give a totally different response due to the appearance of these mutations in the binding pocket.

**1.6. The role of molecular simulation**

In this framework, molecular modelling can be extremely useful supporting the experimentalists and the clinicians in the understanding of these phenomena and of the reasons that drive the loss of affinity of proteins toward certain drugs.

The aim is to create a virtual system that is able to reproduce the real binding event. By evaluating the energetic that drive this complexation between protein and drug we are able to understand how much the binding is favored by nature. Moreover, by analyzing the binding between drugs and mutated kinases, and by comparing the data with the one of the “wild type” case, it is possible to quantify the negative (or positive) effect introduced by the mutation in the stability of the binding site of proteins.

The further step in the analysis of protein-drug binding is to use the modelling techniques as a real “virtual microscope”. In fact, it is possible to manipulate energetic data related to the binding in order to analyze the results with unique precision – simulations can provide a comprehension of this kind of phenomena that is rather unique.
In summary (the details will be explained in the further sections), with simulation we will be able to analyze all the single zone of a certain protein that favors or disfavors the binding with a certain drug and vice versa – a first step in the comprehension of “why?” a therapy works or not and “how?” to modify it in order to reach better results.
2.1. Introduction: the molecular simulation

Molecular modelling has been considered during the last years a fundamental part of the theoretical chemistry. In its early steps the support provided by molecular simulation was hardly something more than movies and graphical representation. However, the level of reliability reached in the development of dedicated force fields and in both simulations and data analysis procedures evidenced the modelling as an additional and fundamental support in branches like molecular and drug design and in the understanding of the molecular behavior.

Basically, the role of molecular simulation is to give “directions” in the research. In further sections, we will show how the modelling can be a real support for scientists. In this chapter, we will introduce and describe the molecular modelling techniques used in this thesis. The procedures adopted in this framework will be described in details starting from the creation of the molecular systems to the most detailed data analysis. Since the basis of the computational methods used in this thesis are common for all the cases of interest, in this section we will present the common procedure, while the single exceptions will be described in the Results & Discussion section dedicated to each case. Part of the presented procedures have been used for the publications related to this thesis.\textsuperscript{28,42,43,44}

2.2. The atomistic scale of simulation

All the studies presented in this thesis regard the atomistic scale of simulation, being the space and time scales in the order of angstroms and nanoseconds respectively. In the field of life science the atomistic level of simulation has been the most widely used and developed in the last years, because it is the one that provide the most exhaustive description of physical-chemical phenomena at an acceptable computational cost.

The atomistic scale of molecular modelling is divided basically in molecular mechanic (MM) and molecular dynamic (MD).
The molecular mechanic (MM) is the branch of the molecular simulation that studies the equilibrium conditions of a certain molecular system. The goal is the determination of the reliable configuration that a certain molecule assumes in given conditions – i.e. equilibrium in Nature. The instrument used to reach this result is the minimization of the total energy of the molecular system. It is worth noting that all the measured energetic and structural quantities are statistic and thus the result is a statistical equilibrium.

In order to analyze the equilibrium conditions of a certain molecule it is necessary to have the minimum information to characterize the whole systems:

1. The position of atoms that compose the molecule in space.
2. The description of how these atoms are bonded together and how “they feel” each other.

For an accurate description of the energetic of a system, it is necessary to determine all the forces that are applied on the atoms that constitute the molecule. Thus, the determination of the total energy of the system is determined as the sum of all the forces that are acting on each atom of the system itself.

The atomistic scale of molecular simulation has a pure classical expression and no quantum term is taken into account, since the electrons of each atom are not explicitly considered. The atoms that constitute a certain molecule are simply represented as balls (representing the atomic nucleus) identified by coordinates and connected each other to compose a certain geometry. While the starting position is described as cartesian x,y,z coordinates in a configuration file, the connection between atoms is a rather more complex issue. The starting point is that molecules follow classical laws in MM and MD, the atoms are simply connected by springs with a certain force constant. These constants depend on the atoms coupled and on the bond type they are able to form. As we will underline in next sections, these terms are described in the force field. The force field is a file that contains a data bank of bond and non-bond parameters for different atoms. There are different force fields dedicated to different branches of applications of molecular modelling (in these study we will use typical force fields dedicated to life science
applications). When some parameters are missing in the force field – i.e. when not all the atoms in the system are present in the force field (like a Silicon atom in a life science dedicated force field) – the software is not able to recognize what this atom is and how it is supposed to bond to each other. However, missing parameters can be calculated by \textit{ab initio} techniques or substituted with experimental values (if available).

Therefore, when the system can described completely, the reaching of the equilibrium is described by equation (2.1).

\[ \frac{\partial E_{\text{pot}}}{\partial r_i} = 0 \]  \hspace{1cm} (2.1)

Where the potential energy $E_{\text{pot}}$ is function of the coordinates $r_i$ of the atoms of the system. Even if molecular mechanic is useful reduce bad energetic configurations that typically occur in the early steps of the simulation, it is considered nothing more than a preliminary step for complex systems, since it does not take care of the temperature, that in MD appears only in the kinetic energy term (Eq. (2.2)) and, thus, where the time evolution is considered.

\[ E_{\text{kin}} = \frac{3}{2} k_B T \]  \hspace{1cm} (2.2)

Molecular dynamic (MD) is simply an iteration in time of molecular mechanic – at each time step of the simulation the system moves toward the equilibrium. The concept of this technique is conceptually rather simple: given a starting configuration for the molecular system of interest (position of atoms), the forces that are acting on each atom are calculated, and thus trajectories can be computed as the evolution at each time step of the system toward the equilibrium. Differently form the MM, in MD an initial set of random velocities are calculated from temperature as in Eq. (2.2) – a starting condition that is defined by temperature. This creates atomic movements – the system moves to reach the equilibrium – and new positions and velocities for all the atom of the system are
calculated at each dynamic step. Obviously, the system fluctuates toward the equilibrium maintaining the total energy \(E_{\text{tot}}\) constant.

\[
E_{\text{tot}} = E_{\text{pot}} + E_{\text{c}} = \text{constant}
\]  

\text{(2.3)}

It means that, if the new velocities are too low or too high to maintain the constancy, at each step they are rescaled in order to be consistent with the temperature of the system and with the \(E_{\text{tot}}\) with the so-called \textit{thermostats} and \textit{velocity scaling algorithms}.

More in details, we said that in MD the electrons are not taken in account in the calculation. Thus, the Schrödinger equation assumes the following form, according to the Born-Oppenheimer approximation:

\[
H\Phi \Phi = E\Phi \Phi
\]  

\text{(2.4)}

This equation loses the explicit description of electrons - the system becomes dependent only on the nucleus position (and velocity) of each atom. All the quantum effects are lost, and Eq. (2.4) can be expressed in classical form as follows:

\[
-\frac{dE_{\text{pot}}(R)}{dR} = m \frac{d^2R}{dt^2}
\]  

\text{(2.5)}

In fact, to solve a MD problem, means to solve across time (at each dynamic time step) an equation like:

\[
a_i(t) = \frac{F_i(t)}{m_i}
\]  

\text{(2.6)}

It is thus necessary to calculate the force \(F(t)\), dependent on the position of each atom in space, from the potential \(V(R)\). This will be done for all the atoms that constitute the system:
\[ F_i = -\nabla V(R_i) \quad \text{for} \quad i = 1, \ldots, n \] (2.7)

The potential \( V(R) \) is, as said, defined by the force field. Once \( F(t) \) is calculated – and thus the acceleration \( a(t) \) is known – by integration it is possible to calculate the new coordinates for all the atoms of the system with the so-called Verlet algorithm that is represented by the following equations:

\[ R_i(t + \Delta t) = R_i(t) + v_i(t)\Delta t + \frac{1}{2} a_i(t)\Delta t^2 \] (2.8)

\[ R_i(t - \Delta t) = R_i(t) - v_i(t)\Delta t + \frac{1}{2} a_i(t)\Delta t^2 \] (2.9)

does not seem to produce results as expected. That can be summed to give:

\[ R_i(t + \Delta t) = 2R_i(t) - R_i(t - \Delta t) + a_i(t)\Delta t^2 \] (2.10)

With this algorithm it is possible to compute the new positions of each atom of the system and thus to calculate the new potential energy, as sum of the potential energetic contribution given by each atom, \( E_{\text{pot}} \). Moreover, it is possible to calculate the new velocities of all the atoms:

\[ v_i(t) = \frac{R_i(t + \Delta t) - R_i(t - \Delta t)}{2\Delta t} \] (2.11)

in order to calculate the new values of the kinetic energy as:

\[ E_{\text{kin},i} = \frac{1}{2} m_i \left[ v_i(t) \right]^2 \] (2.12)
and the constancy of the total energy can be expressed as in equation (2.3). This procedure is iteratively repeated at each time step for the whole length of molecular dynamic simulation.

2.3. Software, force field and parametrization

All the molecular simulation techniques are somehow involved in parametrization problems. As we highlighted in previous sections, the choice and usage of the proper force filed, dedicated to the simulation of a certain class of molecules, is of primary importance for the reliability of the modelling. For instance, a force field that is dedicated to general material science could give completely wrong results for proteins simulations, or in most cases could be unable even to start the calculation.

For all the studies presented in this thesis work the AMBER\textsuperscript{45} (version 9) suite of program was used. AMBER (Assisted Model Building with Energy Refinement) is a family of force fields for molecular dynamics of biomolecules originally developed by the Peter Kollman’s group at the University of California, San Francisco. AMBER is also the name for the simulation package that implements these force fields and it is probably the most famous software for molecular simulation in life science.

The form of the AMBER force field is:

\[
V(R^N) = \sum_{bonds} \frac{1}{2} k_b (l - l_0)^2 + \sum_{angles} k_a (\theta - \theta_0)^2 + \sum_{torions} \frac{1}{2} V_n [1 + \cos(n\phi - \gamma)] +
\]

\[
+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ E_{ij} \left[ \frac{\sigma_i}{(R_{ij})} \right]^{12} - 2 \left( \frac{\sigma_i}{(R_{ij})} \right)^6 \right\} + \frac{q_i q_j}{4\pi\varepsilon_0 R_{ij}} 
\]

(2.13)

The meanings of terms in this equation are:

- First term: the bond term. It represents the energy between covalently bonded atoms. The interaction is modelled as an harmonic (ideal spring) force. This is a
good approximation near the equilibrium bond length, but becomes increasingly poor as atoms separate.

- Second term: the angle term represents the energy due to the geometry of electron orbitals involved in covalent bonding.
- Third term: the torsion term. Represents the energy for twisting a bond depending on bond order (e.g. double bonds) and neighboring bonds or lone pairs of electrons.
- Fourth term: the non-bond term represents the non-bonded energy between all atom pairs as decomposed by van der Waals (first term) and electrostatic (second term of summation) energies.

In Eq. 2.13 all the constant (k_b, k_a, etc.) are unknown. In order to use the AMBER force field, it is necessary to have values for the parameters of the force field (e.g. force constants, equilibrium bond lengths and angles, charges). The AMBER software contains a fairly large number of these parameter sets. Each parameter set has a name, and provides parameters dedicated to certain types of molecules.

- Peptide, protein and nucleic acid parameters are provided by parameter sets with names beginning with "ff". The year of creation follows (i.e. ff95).
- GAFF (Generalized AMBER force field) provides parameters for small organic molecules to facilitate simulations of drugs and small molecule ligands in conjunction with biomolecules.
- The GLYCAM force fields have been developed by Rob Woods for simulating carbohydrates.

In all the calculations presented in this thesis, the ff99 was used, being the most widely used in the study of proteins. All the amino acids forming proteins are parametrized with the constant values present in ff99.
However, ff99 provides a very exhaustive description of proteins and nucleic acids, but parameters for molecules different for the conventional organic ones are missing. As anticipated, the AMBER package can integrate missing parameters of a certain system with the GAFF parametrization. The full list of modules present in the AMBER package are the following:

- **leap** is the tool that is used for preparing files necessary to start the simulation.
- **antechamber** calculates the missing parameters for small organic molecules using GAFF forcefield.
- **sander** (Simulated Annealing with NMR-Derived Energy Restraints) is the central simulation program and provides facilities for energy minimization and molecular dynamics with a wide variety of options.
- **pmemd** is a reimplementation of sander by Bob Duke. It is optimized for parallel processing – it works significantly better than sander when while running on more than 8–16 processors.
- **nmode** calculates normal modes (entropy of the system).
- **ptraj** provides facilities for numerical analysis of simulation results. AMBER does not include visualization capabilities; visualization can be performed with VMD, Chimera or other visualization tools.
- **MM-PBSA** allows for implicit solvent calculations on snapshots from molecular dynamics simulations to evaluate free energies of binding.

By using *ab initio* (or semiempirical) techniques, it is possible to calculate parameters for small molecules that are not originally present in the force field (i.e. small drug molecules). In all the studies described in this thesis, missing GAFF parameters for inhibitor drug molecules were calculated with semiempirical Austin model (AM1) calculation method with the *antechamber* module within AMBER 9 corrected with an *ab initio* term for bonds (BCC bond term). This calculation method is one of the more
precise charge calculation method, and can guarantee results with precision comparable to the one of pure quantum-mechanic calculations allowing a consistent reduction in the computational time.

2.4. Creation of the “wild type” complexes

After this fast description of the software package and of the parametrization method used for complexes in the studies described in this thesis, we are going to present here the procedure that was followed for the creation of the molecular complexes, subject of the simulation studies.

It is worth noting that, in order to obtain results as close as possible to the reality, the virtual system must be created to mimic the real conditions. In our cases, the subjects are proteins and drugs, and the external condition that it is worth to reproduce is the human body. With this in mind we will follow a step by step procedure, from the importation of the file containing the topology of protein-drug complexes, to the reproduction of the correct salt concentration (NaCl) in the surrounding solution.

As default topology files, AMBER 9 is able to read a series of different formats (.pdb, .xyz, etc.). Typically, for protein structure, the format that is most widely used is the .pdb file. The extension .pdb stats for, and is downloadable from, the Protein Data Bank (http://www.rcsb.org/pdb/home/home.do) – a free online data base containing the .pdb files for the most studied protein and organic structures. The file .pdb is a simple coordinate file obtained by X-ray techniques – it contains only the atom numbers and names, the name of the residues to whom these atoms belong and the coordinates of each atom within the structure. Of course, if the molecule represented in the .pdb file is a protein, the residues that form the whole protein structure are amino acids.

```
ATOM      1  N   TYR A 547  -12.320   8.712  -5.526  1.00  20.31           N
ATOM      2  CA  TYR A 547  -13.492   9.637  -5.395  1.00  20.69           C
ATOM      3  C   TYR A 547  -13.372  10.581  -4.200  1.00  20.80           C
ATOM      4  O   TYR A 547  -13.698  11.762  -4.318  1.00  21.18           O
ATOM      5  CB  TYR A 547  -14.795   8.845  -5.276  1.00  20.68           C
ATOM      6  CG  TYR A 547  -15.105   8.017  -6.499  1.00  20.66           C
ATOM      7  CD1 TYR A 547  -15.697   8.591  -7.622  1.00  20.45           C
```
Scheme 2.1. Part of a .pdb file. Columns represent (starting from the second left column) the atom number, name, the residue and chain of belonging. The first three floating point numbers are x, y and z coordinates of each atom expressed in angstroms (Å). The next three columns are the occupancy, temperature factor, and the element name, respectively.

It is worth noting that the first column report the statement ATOM. There could be atoms within the structure with the statement HETATM also – this is used to identify which atoms belong to a residue that common force fields are able to recognize or not. Typically, in fact, HETATM is used to identify a small inhibitor drug inside the .pdb file. This is of course not comprised as a unique residue in any force field, and an independent parametrization will be necessary.
Scheme 2.2. Coordinates .pdb file for the imatinib drug (residue name: STI). The meaning of each column is the same of the previous Scheme.

As it appears evident, .pdb files (almost all of them) do not report Hydrogen atoms. H atoms are added by each molecular simulation software automatically once the .pdb file is loaded.

For the purpose of this thesis, several .pdb files were used, depending on the protein subject of the study and on the inhibitor drug of interest. In this section, the whole
procedure will be maintained as general as possible, since most of the steps that will follow are common for all the cases studied. The single exceptions that belong to single protein complexes were further mentioned and describer in the discussion phase of data related to each case.

Conceptually, to run a simulation it is necessary to obtain from the coordinate .pdb file two different files, whose extensions are .prmtop and .inpcrd. From a .pdb file, in fact, only the positions of atoms and their name are known, but no information about how they are connected and how they feel each other are present. The software reads the .pdb file and interpret it depending on the force field that was chosen for the correct system description. Thus, .prmtop and .inpcrd files are coupled and descript plenty the molecular system. They represent respectively:

1. .inpcrd: the position of atoms that compose the molecule in space (as .pdb files).
2. .prmtop: the values of the constants to put into Eq. (2.12), taken from the force field that was chosen for calculations (bond constants, torsions, atomic partial charges, etc.).

In order to import the .pdb file of the protein-drug complexes inside the AMBER 9 suite of programs and to obtain the respective .prmtop and .inpcrd files, necessary for running the simulations, a preliminary step was required – the creation of missing parameters for the drug (imatinib, dasatinib, etc.). While importing the .pdb complex file, about these small molecules AMBER 9 recognise only positions, but no information about the bond and non-bond constants are present. In fact, in importing a .pdb for both protein and drug molecules, it is obvious to obtain an error due to the impossibility to recognize the drug by the leap module – the module dedicated to the creation of complexes.

In order to obtain the missing parameters for drug molecules, it is necessary to resort to calculation methods that are able to obtain all the parameters for a molecule from its coordinates only. Going down with the simulation scale, there are the quantum
calculations methods (*ab initio*) that, taking into account in the calculation the electrons, are able to give these parameters, independently to the nature of the molecule to be parametrized. It is worth noting however that these calculation methods are based on the explicit solving of the Schrödinger equation and are extremely heavy from the point of view of computational time. Different *ab initio* techniques uses different simplifications to solve the Schrödinger equation – thus they present a different request of CPU time for the resolution. The *antechamber* module of AMBER 9 allows different methodology for calculation of missing parameters (HF, semiempirical methods). In the studies presented in this thesis we used, as said, the AM1-BCC calculation method, able to produce for such small molecules data of similar precision of the most advanced quantum mechanics methods with a notable reduction in the computational time.

The *antechamber* module accept as input file the .pdb file of the molecule that is going to be parametrized and gives as input a file with extension .prepin. The .prepin file list all the calculated parameters (in .GAFF format) for the drug molecule. Moreover, when *antechamber* fails to complete the calculations for all the constants for Eq. (2.13), the *parmchk* utility available in AMBER 9 substitute the missing bond and angle values with the one that he find to be “similar” in the data base of the selected force field. It is worth noting that to successfully pass this phase, the .pdb file of the drug must have the Hydrogen atoms (that, if not originally comprised, have to be added in a preliminary step).

Ones this step is accomplished, it is possible to load into *leap* the .pdb file of the whole complexes. As said, the “.ff” AMBER force field are not “closed”, but they accept the parameters that are calculated with *antechamber*. While the AMBER force fields use caps lock names (i.e. C3) for force field types of atoms, the “GAFF” force field type calculated with antechamber are expressed with small capitals (i.e. c3). The *leap* module is able to accept both force field types without making any confusion and to write a .prmtop file in which the atoms of the complex are identified by both indexes.

```
This is a remark line
molecule.res
STI INT 0
CORRECT OMIT DU BEG
```
Scheme 2.3. Part of the .prepin file for the imatinib drug (residue name: STI). The third column report the force field types for each atom.

remark goes here

MASS

BOND

ANGLE

nb-ca-cp 69.200 122.630 same as ca-ca-nb
h4-ca-cp 48.200 121.090 same as ca-ca-h4
c-a-c3-n3 66.152 111.475 Calculated with empirical approach

DIHED

cp-cp-nb-ca 1 4.800 180.000 2.000 same as X-ca-nb-X
cp-cp-nb-ca 1 4.800 180.000 2.000 same as X-ca-nb-X

IMPROPER

cia-h4-ca-nb 1.1 180.0 2.0 Using default value
cia-h4-ca-nb 1.1 180.0 2.0 Using default value
cia-ca-cp-cp 1.1 180.0 2.0 Using default value
cia-ca-cp-cp 1.1 180.0 2.0 Using default value
ncia-ca-cp-nb 1.1 180.0 2.0 Using default value
nia-ca-ca-nh 1.1 180.0 2.0 Using default value
nia-ca-ca-nh 1.1 180.0 2.0 Using default value
nia-ca-ca-nh 1.1 180.0 2.0 Using default value
nia-ca-ca-nh 1.1 180.0 2.0 Using default value
nia-ca-ca-cp 1.1 180.0 2.0 Using default value
nia-ca-ca-cp 1.1 180.0 2.0 Using default value

NONBON
Scheme 2.4. Part of the .frcmod file for the imatinib drug (residue name: STI). Unknown parameters in .prepin files are here substituted with the closer available inside the force field.

Ones .prepin and .frcmod files are loaded and accepted by the leap module, the residue STI (used here as an example – the procedure is the same for all the drug considered) will be listed between all the available inside the force field, as it would has ever been comprised in the known residues.
Figure 2.1. Screenshots of the text extension of *leap* module (*tleap*). Within the list it is possible to see the presence of the STI residue between the known ones after the loading of the corresponding sti.prepin and sti.frcmod files.

After this step, it was possible to load the .pdb file of the protein-inhibitor correctly. The *leap* screenshot report the total charge of the protein-inhibitor complex.
Figure 2.2. Screenshot of the *tleap* module reporting the total charge of the complex (a). Image of the protein (1T46.pdb file - KIT receptor - in this example) colored per atom (O in red, C in grey, N in blue and S in yellow). Hydrogen atoms are not shown for clarity. The STI molecule (used as example) is colored and highlighted in green.

As reported in Figure 2.2a, the total charge of the complex is -1. Since the imatinib does not carry any charge, this is due to the protein and, more precisely, to some of the residues that constitutes the proteic chain. In fact, Scheme 2.5 shows that not all the amino acids are neutral.
Scheme 2.5. Chemical representation of all the amino acids. At neutral pH, few of them carry a +1 or a -1 charge (they are Arg R, Lys K and Glu E, Asp D respectively).
Materials & Methods

As the complex is loaded in leap without giving errors, and before proceeding to create around the solute the same conditions of the human body, three couples of .prmtop and .inpcrd files were saved – for the protein-drug complex and for the protein and the drug taken separately. This operation will be useful for the energetic analysis and will be commented in further sections.

The next two steps for the creation of the molecular system that will be simulate are:

1. The solvation of the system: the complex solute is immersed in a water box of variable dimensions.
2. Addition of counterions: a suitable number of Na+ and Cl- ions is added in the solution in order to guarantee the system neutrality and to reproduce the correct ionic strength present inside the body (150 mM NaCl).

About these two steps, few comments must be done. The simulations that will be presented in the next sections will be carried out under periodic boundary conditions – the geometry of the water box have to be consistent with this condition that will be expressed as the replication in the three dimension of the same box. This from the calculation point of view is carried out with the condition that all the atoms and molecules that get off the original box from one side, come in again from the opposite. In this way it is guaranteed that the content of the periodic box is always the same. Basically, there are two geometries for the water box that are used for this purpose: the tetrahedric and octahedric geometries. The dimensions can be chosen explicitly, or only the minimum distance between the solute atoms and the side of the box can be defined. In this case, the water buffer must be grater that the cutoff radius that is used for calculations.

In general, the geometry of the water box is chosen dependently on the geometry of the solute, and in order to reduce as much as possible the presence of water molecules in the system. In fact, if the solute present a spherical symmetry, to chose a octahedric water
box allows a consistent saving of “useless” atoms in the system. On the other hand, if the solute is strongly un-spherical and present a geometry that is not isotropic (i.e. DNA strands), to choose an octahedric box (very similar to a soccer ball) will create a final system full of water molecule that are practically useless for the description of the system. In this case, the tetrahedric geometry is more indicated and guarantee more flexibility.

There are different molecular models to describe the water inside AMBER. The most widely used (and the one we used in all our calculations) is the TIP3P water model. This model for water has the peculiarity that the angle between the two Hydrogen atoms is maintained rigid. Thus the water molecules are something “rigid” in the calculation – this is a simplification, but decreases notably the CPU time required to simulate the system (the greatest part of atoms in the system belong to water molecules of the solution).

After the complex is immersed in the water box, it is necessary to add the counterions in the system. Salt ions can affect consistently the ability to move of the solute, influencing the final dynamic trajectories. The correct amount of ions to add to the system are calculated according to the volume of the water box, in order to respect the 150 mM [NaCl] condition. The Na+ and Cl- ions can be added in the system with the leap module – they are inserted in the system following the electrostatic potential surface of the solute. If during the addition of counterions positional conflict (superposition) with water molecules, constituting the box, are verified, the entire water molecule is substitute with the ion that is placed in the water molecule centre of mass. Since the number of water molecules in the box is rather high (high density), almost all the counterions substitute a water molecule. This create a non-uniformity in the box, since one single atom (Na+ or Cl-) is replacing three atoms (H-O-H) – there are single points of lower density. That is why, for instance, during the first period of NPT dynamic (which allows the volume of the water box to change) the box reduces its dimensions in order to adjust the density. In conclusion, in the calculation of how many ions should be added to the system, the fact that each ion will replace an entire water molecule and that the final
equilibrated volume will be smaller than the one of the initial box constructed by *leap* must be taken into account.

Figure 2.3. The system KIT “wild type” in complex with imatinib. The protein is represented as a sold ribbon with helix colored in red and sheets in blue. The imatinib molecule is colored in yellow. Cl- and Na+ are represented as green and purple spheres respectively. Water is colored in cyan and Hydrogen atoms are omitted for clarity.
Once the system is completed it is possible to save the .prmtop and .inpcrd files for the whole system – these are files that characterize the system that is going to be simulated.

### 2.5. Creation of mutated complexes

In the previous section, we described how to construct the “wild type” complex – in absence of mutations in the protein chain. However, the main aim of the studies performed and exposed in this thesis is to evaluate and analyze the effect that amino acid mutations have on the inhibition effect of the drug.

In the introductive section we already described the mutated kinases as result of mutations that arise into the genes that express these proteins. The causes of these genetic mutations are not well known but the mutated kinases, expressed by the mutated genes lose the control of external factor and present an un-controlled tyrosine-kinase activity. It is worth mentioning also that amino acid mutations can arise during the treatment – these are the so called “secondary” mutations. Moreover, these genetic anomalies can originate different kind of mutations:

1. Point mutation: substitution of a certain amino acid with another one in a precise position of the protein chain
2. Amino acidic deletion: one or more amino acids, present in the “wild type” protein, can be simply absent in the mutated kinase
3. Amino acidic duplication: one or more amino acids, present in the “wild type” protein, can be duplicated in the mutated structure

Actually, in the studies presented in this thesis, only the first two families will be analyzed. The third one, the rarest case, present not trivial problems in the modelling due to the high atomic density of the protein structure. In fact, while the substitution (point mutations) or the deletion (amino acid deletions) of something on the “wild type” structure is a procedure that can be easily accomplished, the addition of new material
inside the protein chain is practically impossible without a complete folding procedure
due to the unavoidable geometrical conflict and superposition that would be generated.

Thus, in order to analyze how much a certain drug is able to inhibit a mutated kinase as
well as the “wild type”, it is necessary to model also the mutated proteins and to insert the
mutation inside the naive protein.

In this section we will briefly describe how we introduced point mutations and amino
acid deletions inside the protein chain. To this purpose we used the biopolymer module
available inside the Insight II software (version 2000.1) by Accelrys Inc. (San Diego,
USA).

2.5.1. Point mutations

Using the biopolymer module, it was possible to replace a single amino acid within the
protein chain with another one. The Insight II software automatically compute one (or
more if available) configuration for the new amino acid in order to avoid superposition
with the atoms of the surrounding and in order to reach the minimum energy. If more
than one configuration is available, biopolymer shows all the possible configurations with
the respective energies – obviously, in all these cases, the configuration with minimum
energy was chosen and the mutated protein was saved as a new .pdb file.

2.5.2. Amino acidic deletions

To introduce a deletion along the amino acidic chain is a more immediate operation. It
is enough to delete the residues directly from the .pdb text file, or, using a graphical
extension (biopolymer module), to delete them explicitly and to create a new peptide
bond between the backbone atoms of the desired amino acids. The result will be a very
long peptide bond between two residues that in the “wild type” structure were separated.
However, this long bond does not create problems as during the rising of the temperature
(to go from 0K to 300K at regime) the entire proteic structure will undergo to
readjustment and the solute will reach a new adapted configuration.

It is evident from Figure 2.4. that, while a point mutation provoke only adjustments in
the secondary structure of the protein, a consistent deletion can provoke drastic
distortions and final configurations that are considerably distant from the “wild type” kinase. Thus, it is perfectly clear that the huger the deletion, the more consistent the simplification and the risk that the final mutated protein conformation in the model could be completely different from the real one.

Figure 2.4. Examples of point mutation (a: L576P) and amino acidic deletion (b: del570-576) and their effect on the structure. In the “wild type” (left) complexes the mutated (or deleted) residues are colored in red. It is evident how different mutations result in different distortions within the mutated complexes (right).
Once the mutation has been inserted, the same steps of solvation and addition of counterions described in the previous section followed – a .prmtop and an .inpcrd file were obtained also for the mutated case, ready to be simulated.

2.6. The simulation procedure

With the two files containing parameters (.prmtop) and starting configurations (.inpcrd) we had everything is needed as a starting point to run a molecular simulation. Together with the subject of the simulation, what is needed is an engine for calculations and an input file. As engine for calculations the Sander.MPI (Simulated Annealing with NMR-Derived Energy Restraints) and pmemd modules of AMBER were used. Both modules are parallel calculation tools, but while Sander.MPI allows many facilities and tools for energy minimization, molecular dynamic, restraints, etc., pmemd presents only few options. In fact, with pmemd, only periodic boundary condition simulations without restraints applied on the molecule are allowed, but a better scaling during the parallelization is possible – it was used in all the cases studied in the equilibration and production phase of our simulations.

At the first step our system is in the configuration described by the .inpcrd file – as said, the density of the solution in the box is not uniform in space due to the presence of single ions that replaced water molecules and the temperature is still 0K. In all the cases presented in this thesis, the simulations phases can be summarized as follows:

1. Minimization: a preliminary phase, only necessary to remove bad contacts between atoms.
2. Constrained dynamic: a phase in which the solute is constrained in its starting position – during this phase the simulation temperature is reached, and start the solution density optimization.
3. Equilibrium dynamic: without any restraint – it is the phase in which the system reach the equilibrium (NPT).
4. Production dynamic: without restraints in NVT condition – the trajectories of this phase will be further processed and analyzed.

In detail, the molecular systems were first minimized for 500 steps using Sander.MPI applying a restraint on all the atoms of the solute with elastic constant of 2 N Å\(^{-1}\) in order to let move all the atoms belonging to the solution by first. After that, a second minimization identical to the first, but in absence of restraints on the solute was run to remove bad contacts between atoms of the complex. After the minimization part, 50000 steps with a time step of 1 femtosecond (fs) of NVT dynamic was run to carry the initial temperature of 0K to the production one of 300K. To this step for temperature relaxation, other 50 picoseconds (ps) of molecular dynamic in NPT conditions (time step 1 fs) at the constant temperature of 300K at the pressure of 1 atm was run in order to relax the density of the solvent. Both these two first dynamic steps were conducted with the solute maintained restrained on the starting position applying an elastic force with a 2 N Å\(^{-1}\) elastic constant. After this step, the equilibrium phase was conducted in NPT conditions removing all the restraints. In this phase, the solute is supposed to reach the equilibrium. More precisely, this phase can take from few to tens of nanoseconds, depending on the conformation of the solute. This step is conducted using the pmemd module at the constant temperature of 300K and, using the SHAKE\textsuperscript{48} algorithm to constrain all the bonds involving hydrogen atoms, it was possible to enlarge the time step to 2 fs. We preferred to use in the previous steps 1 fs time step in order to avoid too large deviations due to the rising of temperature and to the high fluctuations of the molecules belonging to the solvent. The equilibrium phase, thus, has not a pre-defined length – it depends on how long the reaching of the equilibrium takes to the complex. Therefore, during the simulation, few quantities are monitored to check the status of the equilibrium of the system. The quantities that usually are monitored to check the status of the system are:

1. The potential energy \(E_{\text{pot}}\): strictly correlated to the total energy of the system \(E_{\text{tot}}\). The energy of the system usually converges almost immediately (i.e. until the first picoseconds of the equilibrium phase).
2. The root mean square deviation (RMSD) of the atoms of the backbone

While the first parameter tends reach an equilibrium immediately (in correspondence of the stable reaching of the 300K temperature), the second takes a longer.

\[
RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \delta_i^2}
\]  

(2.14)

Where \( \delta \) is the distance between \( N \) pairs of equivalent heavy atoms from their starting position (in this study the atoms belonging to the backbone of the protein). If the equilibrium is reached, both the curves of potential energy and RMSD should be horizontal. On the other hand, if the RMSD has not reached the equilibrium (it is growing) despite a constancy of the energetic of the system, it means that the complex, in order to maintain the constancy of the energy, is vibrating stronger and stronger at each dynamic step - thus, while the energetic equilibrium is reached, the vibrational one has not.

As reported by Figure 2.5, the energies converge sooner than RMSD – the real equilibrium is reached after \(~6\) ns in this case, thus this is the starting point from which the data collection can be considered as reliable. In conclusion, the equilibrium phase has not a pre-defined length, but it depends on the time required to the solute to reach the equilibrium. It is also worth noting that a good oscillation for RMSD is between 1 and 2 Å around the equilibrium position – in this case, if we zoom on the terminal part of the RMSD curve (last 2-4 ns) the fluctuation of atoms around their equilibrium position is really below these values.

After the system has completely reached the equilibrium, the production phase can start. It is conducted for 2 nanoseconds in NVT conditions using the pmemd module of AMBER. The other parameters are the same of the equilibrium phase.

In all these simulations step a cutoff was used. The role of the cutoff is to exclude from an explicit non-bond calculation all the atoms that are not inside a sphere defined by the cutoff radius (\( R_{\text{cutoff}} \)). This is done for shake of CPU time saving. Since the cutoff value
can vary depending on the single systems, the individual parameters and details for each case studied and presented in this thesis will be reported in the Result & Discussion section, in the proper paragraph dedicated to each case.

Figure 2.5. Examples of $E_{\text{pot}}$, $E_{\text{tot}}$ and RMSD profiles of a protein-drug complex (KIT “wild type”-imatinib). The time scale is expressed in nanoseconds, the energies in kcal mol$^{-1}$ and the RMSD in angstroms (Å). The last 2 ns of these profiles belong to the production phase of the dynamic.

However, since all the atoms do not enter in the calculation explicitly, a corrective term (an external constant) must be inserted, in order to correct the error given by the exclusion of part of the atoms from calculation. This is done using the Particle Mesh Ewald (PME) summation. Briefly, since the electrostatic interaction can be divided in short- and long- range effects, this methodology is based on the following concept:

$$\varphi(r) = \varphi_{sr}(r) + \varphi_{lr}(r)$$  \hspace{1cm} (2.15)

Where the non-bond interaction potential $\varphi(r)$ is given by the sum of a short range ($\varphi_{sr}(r)$) and a long range ($\varphi_{lr}(r)$) term. In this conception, the first one is the one calculated explicitly in the real space (the one defined by the cutoff radius), and the
second one is calculated with a Fourier summation and converges rapidly in the calculation.

Figure 2.6. Equilibrated dynamic snapshot of KIT-dasatinib complex. The receptor is represented as a solid ribbon and the dasatinib molecule as a yellow stick. Cl- and Na+ ions are represented as green and purple spheres. Water is colored in cyan and Hydrogen atoms are omitted for clarity.

These are the basics of the simulations that were conducted and are presented in this thesis. In conclusion the total length of each simulation was ~10 ns (depending on each system, but the complexes presented in this thesis are rather similar).

The single simulation details (number of atoms of the system, number of water molecules, cutoff used, etc.) concerning each single case studied will be reported in the Result & Discussion section, in the paragraph dedicated to each single system simulated.
2.7. Affinity energy analysis (MM-PBSA)

In the previous section, we described how the simulations were conducted. But once the simulations are finished, the output files must be processed in order to obtain relevant results. At the end of the simulations the output files we have obtained are:

- File .out: the output file. It reports all detail about the energetic of the system at each step of the simulation. It also list all the parameters listed in the input files.
- File .rst: file restart. It is equivalent to the .inpcrd – it is a coordinate file that is written after a given time interval. In dynamic regime, differently from the .inpcrd file, it reports also the velocities of the atoms.
- File .mdcrd: the trajectory file. These files have huge dimensions and reports the “history” of the simulation for each atom of the system.

The last file is the most important one for our purposes – we are going to process the dynamic trajectory files to perform a MM-PBSA analysis on the complex in order to evaluate, at each step of the simulation, the affinity energy that exist between ligand and receptor.

The MM-PBSA\(^{50}\) methodology (Molecular Mechanic – Poisson Boltzmann Solvent Accessible area) is a tool that is implemented in AMBER to calculate the binding free energy for the association of two molecules (in this case a protein and an inhibitor drug) in solution. The main aim of the MM-PBSA method is to calculate the free energy difference between two states which most often represent the bound and unbound state of two solvated molecules or to compare the free energy of two different solvated conformations of the same molecule.

\[
[L]_{aq} + [R]_{aq} = [L^* R^*]_{aq},
\] (2.16)

Ideally, the free energy of binding for the association of a ligand (L: drug) and a receptor (R: protein) can be expressed by the following representation:
In the simulation of these solvated states the majority of the energy contributions would come from solvent-solvent interactions and the fluctuations in total energy would be an order of magnitude larger than binding energy of the two molecules of interest. Thus the calculation would take an enormous amount of CPU time to converge. In order to simplify, the MM-PBSA method is based on the following thermodynamic cycle:

Figure 2.8. Thermodynamic cycle that explains how the $\Delta G_{\text{bind,solv}}$ of the solvated state can be calculated as function of the $\Delta G_{\text{bind,vacuum}}$ of calculated in vacuum.$^{51}$

From Figure 2.7, the binding free energy $\Delta G_{\text{bind,solv}}$ can be calculated by:
\[ \Delta G_{\text{bind,solv}} = \Delta G_{\text{bind,vacuum}} + \Delta G_{\text{solv,complex}} - (\Delta G_{\text{solv,ligand}} + \Delta G_{\text{solv,receptor}}) \]  

That is as to add to the free energy of binding calculated in vacuo a correction for solvation:

\[ \Delta G_{\text{bind,solv}} = \Delta G_{\text{bind,vacuum}} + \Delta G_{\text{solv}} \]  

In the MM-PBSA approach the different contributions to the binding free energy above are calculated in various ways: solvation free energies are calculated by either solving the linearised Poisson Boltzman equation for each of the three states (this gives the electrostatic contribution to the solvation free energy) and adding an empirical term for hydrophobic contributions:

\[ \Delta G_{\text{solv}} = G_{EL,\epsilon=80} - G_{EL,\epsilon=3} + \Delta G_{\text{NP}} \]  

\( \Delta G_{\text{bind,vacuum}} \) is obtained by calculating the average interaction energy between receptor and ligand and taking the entropy change upon binding into account:

\[ \Delta G_{\text{bind,vacuum}} = \Delta E_{\text{MM}} - T\Delta S \]  

and thus Eq. (2.18) can be written as:

\[ \Delta G_{\text{bind,solv}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S \]

The entropy contribution can be calculated by performing a normal mode analysis\(^{52}\) on the three species. The normal mode analysis calculations are computationally expensive and are the one that introduce in the results the larger margin of error and the higher uncertainty in the result.
It is demonstrated that the entropic term \(-T\Delta S\) for kinases, like the ones we treat in this study inhibited by a small drug (like imatinib, dasatinib, etc.), is rather constant and \(-17\) kcal mol\(^{-1}\). In case of point mutations this number does not change a lot, since the difference between the two complexes is only of few atoms and can be considered (as we will see) as a constant. On the other hand, this value tends to change in presence of amino acidic deletions. Here the lack of a certain number of amino acids can induce important modifications in the ability to move of the entire structure blocking many degrees of freedom – this can modify the entropic cost of binding.

\[
\Delta E = E_{\text{complex}} - (E_{\text{receptor}} + E_{\text{ligand}})
\]  

(2.22)

At each step of molecular dynamic the energetic quantities of the previous equations are of the form of the \(\Delta E\) of Eq. (2.22). These delta-energetics measure the tendency that ligand and receptor have in the human body to form a stable complex rather than to maintain two individual and separated configurations – the more negative the \(\Delta E\), the more stable the complex (it means that in certain condition the complex has lower energy than the two of protein and drug taken separately and, thus, its formation is favored).

The methodology that is adopted in this thesis for the calculation of each energetic term of Eq. (2.22) is the “single trajectory” technique. In fact, starting from the assumption that the conformational change of protein and drug will not be heavily affected due to the binding, the single conformations of protein, drug and complex for which the energies will be calculated at each steps will be taken from the single trajectory of the system we simulated. Practically, from the single dynamic trajectory of all the periodic box, we define “what” is complex and “what” is ligand and receptor respectively – this is what is needed for the calculation of each energetic component for the calculation of the free energy of binding \(\Delta G_{\text{bind}}\).

This procedure is of course faster than a “many trajectories” approach, because it allows to run a single simulation instead of three.

However, in such simplified approach, the presence of water and ions is no more taken into account into the energetic calculations but only in the effect on the trajectory of these
molecules along the dynamic. As said, the polar component of solvation $\Delta G_{\text{EL}}$ is calculated with the Poisson Boltzmann method\textsuperscript{53} and the non polar component $\Delta G_{\text{NP}}$ is calculated as:

$$\Delta G_{\text{NP}} = \gamma (\text{SASA}) + \beta$$ \hspace{1cm} (2.23)

in which $\gamma = 0.00542 \text{ kcal/Å}^2$, $\beta = 0.92 \text{ kcal/mol}$, and SASA is the solvent-accessible surface estimated with the MSMS program.\textsuperscript{54}

In practice, we considered for the calculation of the free energy of binding the last 2 ns of the equilibrated phase of the single molecular dynamic trajectory (the production phase). A certain number (the precise number can vary with the system, being related to the dimensions of molecules – precise details on the single cases will be provided in the dedicated Result and discussion section) of snapshots were obtained from the single dynamic trajectory using the mm_pbsa.pl script of AMBER for the complex, and the protein and drug unbound. After this, the same mm_pbsa.pl script was used in order to compute the free energy of binding of each complex of interest in this thesis. simulated.

2.8. Per residue decomposition analysis

In previous section we explained the methodology used for the calculation of the free energy of binding $\Delta G_{\text{bind}}$. The effectiveness of a certain inhibitor drug is defined by the IC50 – the amount of drug to put into the solution in order to inhibit the activity of the kinase of the 50%. The correspondence between IC50 and the free energy of binding is defined as follows:

$$\Delta G_{\text{bind}} = -RT \ln(\text{IC50})$$ \hspace{1cm} (2.24)

Thus, until this point, we were able to produce results consistent with experimental measurement, indicating that the procedures and the molecular models used for the cases faced in this thesis are reliable. Moreover, from the qualitative point of view we were
able to predict if a drug can inhibit more or less effectively a certain kinase, or whether a precise mutation in a given position of the amino acidic chain can somehow affect the stability of the complexation.

In this section we introduce a data analysis that can give interesting answers to the question “why?”. The single cases in which this has been applied will be mentioned in the proper Result and discussion section.

The concept is rather simple: it is possible to split the total interaction energy (Eq. (2.20)) that exists between ligand and receptor on all the residues constituting the protein:

$$\Delta E = \sum_{i=1}^{n} \Delta E_i$$

(2.25)

By calculating each $\Delta E_i$ it is possible the part that each amino acid has in creating a stable protein-drug complex – the more negative the $\Delta E_i$, the stronger is the favorable action of the residue in stabilizing the complex. In details, the gas-phase in vacuo energies ($E_{gas}$) for each residue are composed of electrostatic and van der Waals interaction contributions ($E_{ele}$ and $E_{vdW}$, respectively) according to equation below.

$$\Delta E_{i,gas} = \Delta E_{i,vdW} + \Delta E_{i,ele}$$

(2.26)

The in vacuo gas-phase energy for each residue ($\Delta E_{i,gas}$) is then corrected according to solvation to give the total energy $\Delta E_{i,tot}$.

Since the $mm_pbsa.pl$ script of AMBER 9 does not support the Poisson-Boltzmann$^{55}$ solvation method for residue energy decomposition, as described in the previous section for the calculation of the free energy of binding, the Generalized Born method was therefore used to correct the gas-phase energies for solvation.

The result is an enthalpy division between all the residues constituting the protein, and gives precious indications in which are the residues that are crucial in binding – the residue that most interact favorably with the inhibitor molecule.
Figure 2.9. Superposition of the “wild type” and mutated L576P KIT protein. The residues that are mostly interacting with imatinib are colored in blue and orange for “wild type” and L576P complexes. The mutation is highlighted by the green oval and the mutating residues and the imatinib drug are represented in stick. It is evident how the mutation in a certain position of the binding site can provoke an important modification of the amino acidic structure and in the configuration assumed by the inhibitor due to a “domino effect”. The protein is represented as a grey shadow.

Moreover, the same procedure is done for the mutated kinases – by decomposing with this per-residue criterion the total interaction energy $\Delta E_{i,\text{tot}}$ it is possible to compare the interaction energies of each amino acid toward the inhibitor in the two complexes “wild type” and mutated.
\[ \Delta \Delta E_{i,\text{tot}} = \Delta E_{i,\text{tot}} - \text{"wild type"} - \Delta E_{i,\text{tot}} - \text{mutated} \] (2.27)

The \( \Delta \Delta E_{i,\text{tot}} \) value represents the effect of the mutation on the affinity between a precise residue and the drug. In fact, if \( \Delta \Delta E_{i,\text{tot}} < 0 \) it indicates a negative effect of the mutation on the attraction ability of the single residue toward the drug – the \( \Delta E_{i,\text{tot}} - \text{mutated} \) is less negative (favorable) than \( \Delta E_{i,\text{tot}} - \text{"wild type"} \), meaning that this precise residue tends to stabilize more the binding in the naive conformation of the protein than in presence of a given mutation along the amino acidic chain. It is worth noting that of course the residues that will report the most evident getting worse due to the presence of the mutation will be the ones in close proximity to the mutation itself.

This energetic analysis can be rather useful because it is able to provide unique details in the binding between protein and inhibitor drugs. Such details are impossible to obtain with other experimental techniques, and evidence this approach and the modelling in general as an extremely powerful instrument for the exploration and the analysis of the potentiality of different small inhibitor molecules.
3.1. Preface

In this chapter we are going to present in details the results achieved during the PhD activity. The result and discussion will be divided per-argument and details regarding each case studied will be provided and commented in this section. All the simulation work presented in this chapter has been carried out in collaboration with experimental partners (Istituto Tumori of Milan and M.D. Anderson Cancer Center in Houston, Texas). Details regarding the experimental techniques used by these partners are available in the full published articles – in this chapter only the computational insights will be commented. This section highlights the important support that molecular modelling is able to provide to experimentalists due to its extreme flexibility and to the possibility to reproduce and analyze ideal conditions that are not reproducible in the experiments.
3.2. The T670I mutation in KIT

The first case presented in this thesis is the study of the mutation T670I in the KIT receptor. This mutation is interesting *per se* because it has an homologous also in the BCR-ABL and in the PDGFRA proteins. The study was conducted in collaboration with the Istituto Tumori of Milan and concerned the understanding of why, in all the imatinib resistant cases, the T amino acid is always substituted by a I.


3.2.1. KIT T670, Bcr-Abl T315I and PDGFRA T674I

As already underlined, different pathologies can have a common genetic origin. Cells that are interested by the tumoral diseases express genes that, because of not well known causes, can mutate during the life-time of the patient. These genes synthesise proteins or receptor whose activity loses the control by grow factors – they are constitutively always active and phosphorylated.

Chronic myeloid leukemia (CML), gastrointestinal stromal tumor (GIST), and idiopathic hypereosinophilic syndrome (HES) represent three different cancers that are associated with pathological deregulation of specific tyrosine kinases: BCR-ABL, KIT, and PDGFRA, respectively. The nature and origin of these three proteic structures has been already described in Chapter 1. However, the most important characteristic of each of these proteins remains its constitutively activated, tyrosine kinase domain. Therapeutic inhibition of these tyrosine kinases by imatinib mesylate confers clinical benefits and substantial anticancer activity in most patients with the above-mentioned diseases. Imatinib is an orally 2-phenylaminopyrimidine compound that was designed to inhibit ATP binding to tyrosine kinase domains in BCR-ABL and was subsequently found to be a successful inhibitor of KIT and PDGFRA by blocking the activation of kinase-mediated intracellular signal transduction pathways. However, kinase inhibition by imatinib has not been uniformly successful. Imatinib resistance in GIST usually
results from selection for the preferential growth of cells that have acquired secondary point mutations in the part of the gene that encodes the KIT kinase domain.\textsuperscript{63-68} Among these imatinib-resistant KIT mutants, only one shared amino acid substitution affecting the ATP-binding pocket of KIT and PDGFRA has been reported, that is, Thr670Ile in KIT and Thr674Ile in PDGFRA.\textsuperscript{58,67} The substitution of both amino acids has been reported to induce a change in the conformation of the ADP binding site,\textsuperscript{70,71} which results in a dramatically lower affinity of the mutant receptors for imatinib.\textsuperscript{64-67,69,70}

Interestingly, an alignment of the KIT and PDGFRA primary sequences revealed that Thr670 in KIT and Thr674 in PDGFRA occur at positions homologous to Thr315 in BCR-ABL, a residue reported to be mutated in imatinib-resistant CML patients.\textsuperscript{72-76} Importantly, Figure 3.1 represents the superposition of the secondary structure of the binding site of BCR-ABL, KIT and PDGFRA. It is immediate to note the extreme superposition of the three kinases in the binding zone and the homologous positions of the threonine residues T315, T670 and T674 in BCR-ABL, KIT and PDGFRA respectively.

![Figure 3.1. Superposition of the secondary structures of BCR-ABL (yellow), KIT (pink) and PDGFRA (green). The high level of binding site superposition is highlighted by the mutual positions of the threonine residue – position 315 (yellow), 670 (pink) and 674](image-url)
(green) for BCR-ABL, KIT and PDGFRA respectively – that is almost superimposed. Atoms and bonds of proteins, ions and water molecules are omitted for clarity.

During the synthesis of a certain protein, each of its constitutive amino acids is originated by a triplet of DNA bases – Figure 3.2 reports all the possible combination of different amino acids that can substitute a threonine due to the permutation of each base.

<table>
<thead>
<tr>
<th>1&lt;sup&gt;st&lt;/sup&gt; base</th>
<th>ACA</th>
<th>ACC</th>
<th>ACG</th>
<th>ACT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2&lt;sup&gt;nd&lt;/sup&gt; base</th>
<th>ACA</th>
<th>AGA</th>
<th>ATA</th>
<th>AAA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T</td>
<td>R</td>
<td>I</td>
<td>K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3&lt;sup&gt;rd&lt;/sup&gt; base</th>
<th>ACA</th>
<th>GCA</th>
<th>TCA</th>
<th>CCA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T</td>
<td>A</td>
<td>S</td>
<td>P</td>
</tr>
</tbody>
</table>

Figure 3.2. Possible combinations of amino acids substituting threonine (T: blue) due to the modifications of each base. The possible mutations (R, I, K, A, S and P) are colored in red.

As it is possible to understand by Figure 3.2, a threonine (T) can be replaced with equal probability by alanine (A), proline (P), serine (S), isoleucine (I), arginine (R) and lysine (K). Surprisingly, in all imatinib-resistant cases involving any of these tyrosine kinases, threonine was substituted exclusively by isoleucine at this position. To understand why, in the presence of imatinib, only isoleucine is naturally selected to replace threonine over all possible amino acid substitutions that can be generated by point mutation at the
Thr670 codon, we introduced all the other possible mutations within the KIT Δ559 structure, in order to understand if there is a thermodynamic motivation beyond the selection of the I amino acid.

3.2.2. T670X: the results

As said, the “wild type” Δ559 KIT-imatinib complex was created (starting from the 1T46.pd file) and simulated following the procedure described in the Material & Method section.

Figure 3.3. Mutant residues in position 670 within the binding site of the KIT receptor complexated with imatinib. The secondary structure of KIT is represented as a dark shadow, as well as the imatinib molecule. In position 670, the “wild type” threonine is
colored per atoms, while the mutations R, I, K, A, S and P are colored in yellow, red, green pink, blue and orange respectively.

The same procedure was adopted also for the mutated cases. We created a mutated complex for each of the possible amino acid substitutes (S,P,K,A,R and I) using the Biopolymer module of Insight II. Calculations were run using the Sander.MPI and the pmemd modules within the AMBER 9 software. Figure 3.3 represents all the mutant residues in position 670.

The equilibrated trajectories for all the complexes were further analyzed using the mm_pbsa.pl module of AMBER 9 in order to obtain the free energies of binding for all the complexes simulated. We used the MM-PBSA approach\textsuperscript{50} to compute the affinity energies as composed by an enthalpic and an entropic term (see Section 2.7 for details). The thermodynamic results are reported in Table 3.1 – all the energies are expressed in kcal mol\textsuperscript{-1}.

Table 3.1. Free energies of binding and energetic components for KIT-imatinib complexes expressed in kcal mol\textsuperscript{-1}.

<table>
<thead>
<tr>
<th>Energy</th>
<th>KIT + imatinib</th>
</tr>
</thead>
<tbody>
<tr>
<td>(kcal mol\textsuperscript{-1})</td>
<td>WT</td>
</tr>
<tr>
<td>ΔE\textsubscript{EL}</td>
<td>-56.0±0.2</td>
</tr>
<tr>
<td>ΔE\textsubscript{vdW}</td>
<td>-72.1±0.3</td>
</tr>
<tr>
<td>ΔE\textsubscript{MM}</td>
<td>-128.1</td>
</tr>
<tr>
<td>ΔG\textsubscript{PB}</td>
<td>107.2±0.3</td>
</tr>
<tr>
<td>ΔG\textsubscript{NP}</td>
<td>-6.7±0.1</td>
</tr>
<tr>
<td>ΔG\textsubscript{solv}</td>
<td>100.5</td>
</tr>
<tr>
<td>-TΔS\textsubscript{MM}</td>
<td>17.4</td>
</tr>
<tr>
<td>ΔG\textsubscript{bind}</td>
<td>-10.2±0.2</td>
</tr>
<tr>
<td>ΔΔG\textsubscript{bind}</td>
<td>-3.8</td>
</tr>
</tbody>
</table>
We calculated the $\Delta G_{\text{bind}}$ as composed by the energetic components reported in the Table 3.1. By comparing the free energies of binding of the WT complex with the mutated ones it was possible to quantify the effect the mutation has on the thermodynamic stability of the binding site. In particular, we calculated $\Delta \Delta G_{\text{bind}}$ as the difference between mutated and WT complex – the more negative the $\Delta \Delta G_{\text{bind}}$, the more crucial the role played by the mutation in destabilizing the binding with imatinib.

Data in Table 3.1 show that the negative effect of the substitution of the T670 with an I is in the same order of the one reported also by P and R. On the other hand, A, S and K show a lower getting worse in the thermodynamics of the binding site – T670A, T670S and T670K are successfully inhibited by imatinib.

These results are consistent with experimental evidences reported by Istituto Tumori of Milan. The WT $\Delta 559$ KIT is successfully inhibited by imatinib, similarly to T670A, 670S and T670K -$\Delta 559$ KIT complexes. On the other hand $\Delta 559$- T670P and T670R KIT together with, the Thr670Ile mutant in the $\Delta 559$ background continued to be strongly phosphorylated.

Of all the systems tested, Thr670Ile, $\Delta 559$ KIT was least sensitive (ie, the most resistant) to inhibition of phosphorylation by imatinib together with Thr670Arg and Thr670Pro.

Starting from the 1PKG.pdb file, we created in a second step the same WT $\Delta 559$ and mutated $\Delta 559$ complexes for the KIT in complex with ADP. We conducted simulations and performed free energies of binding analyses following the same procedure described previously. The results are reported in Table 3.2 as expressed in kcal mol$^{-1}$.

As it is evident from Table 3.2, the KIT protein that contained the Thr670Ile substitution in the $\Delta 559$ background was strongly auto-phosphorylated, similarly to its Thr670, $\Delta 559$/KIT parent. Interestingly, $\Delta 559$/KIT proteins containing the other substitutions at position 670 were expressed, but only weakly phosphorylated (Thr670Ala, Thr670Ser, and Thr670Lys), very weakly phosphorylated (Thr670Pro), or not phosphorylated at all (Thr670Arg).
Table 3.2. Free energies of binding and energetic components for KIT-ADP complexes expressed in kcal mol\(^{-1}\).

<table>
<thead>
<tr>
<th>Energy (kcal mol(^{-1}))</th>
<th>KIT + ADP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WT</td>
</tr>
<tr>
<td>(\Delta E_{el})</td>
<td>-81.3±0.5</td>
</tr>
<tr>
<td>(\Delta E_{vdW})</td>
<td>-82.1±0.3</td>
</tr>
<tr>
<td>(\Delta E_{MM})</td>
<td>-163.4</td>
</tr>
<tr>
<td>(\Delta G_{PB})</td>
<td>130.2±0.2</td>
</tr>
<tr>
<td>(\Delta G_{NP})</td>
<td>-8.6±0.1</td>
</tr>
<tr>
<td>(\Delta G_{solv})</td>
<td>121.6</td>
</tr>
<tr>
<td>(-T\Delta S_{MM})</td>
<td>17.6</td>
</tr>
<tr>
<td>(\Delta G_{bind})</td>
<td>-24.2±0.9</td>
</tr>
<tr>
<td>(\Delta \Delta G_{bind})</td>
<td>-</td>
</tr>
</tbody>
</table>

The same trend was also observed in the clinical evidences reported by our experimental partners. The degree of phosphorylation of all mutants can be ranked as follows: Thr670Ile >> Thr670Ser > Thr670Ala > Thr670Lys > Thr670Pro > Thr670Arg. We concluded that all potential secondary point mutations that caused amino acid substitutions in the ATP pocket of KIT except Thr670Ile were able to attenuate the activating effect of the \(\Delta 559\) deletion in the juxtamembrane domain of the KIT receptor.

We performed a more deep energetic-structural analysis to understand the direct effect of each mutation in destabilizing the binding with ADP and imatinib. The protein-target interaction energies were decomposed on a per-residue base in order to:

1. Identify the residues that mostly stabilize the binding of the ADP with the binding site.
2. Evaluate how the mutation in position 670 affects the binding, by destabilizing the binding action of each of these most interacting residues.
The results are reported in Table 3.3. The non-bond interaction energies of each of the amino acids that are most interacting with ADP are reported as composed by a vand der Waals ($\Delta E_{vdW}$) and an electrostatic term ($\Delta E_{EL}$). It is evident how in the complexation with ADP, the T670I mutation does not play an important role against these residues – the values are practically the same of the WT complex. T670R on the other hand provokes a consistent loss of interactions by these crucial residues – both $\Delta E_{vdW}$ and $\Delta E_{EL}$ values show a less favorable trend in stabilizing ADP.

Table 3.3. Interaction energies associated with intermolecular interactions between ATP and various amino acid positions within wild-type, Thr670Ile and Thr670Arg mutant KITs.*

<table>
<thead>
<tr>
<th>Residue</th>
<th>WT $\Delta E_{vdW}$ kcal mol$^{-1}$</th>
<th>WT $\Delta E_{EL}$ kcal mol$^{-1}$</th>
<th>Thr670Ile $\Delta E_{vdW}$ kcal mol$^{-1}$</th>
<th>Thr670Ile $\Delta E_{EL}$ kcal mol$^{-1}$</th>
<th>Thr670Arg $\Delta E_{vdW}$ kcal mol$^{-1}$</th>
<th>Thr670Arg $\Delta E_{EL}$ kcal mol$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu640</td>
<td>-11.6</td>
<td>-4.7</td>
<td>-11.8</td>
<td>-4.6</td>
<td>-9.8</td>
<td>-2.3</td>
</tr>
<tr>
<td>Glu671</td>
<td>-2.9</td>
<td>-4.2</td>
<td>-3.0</td>
<td>-4.00</td>
<td>-2.8</td>
<td>-3.5</td>
</tr>
<tr>
<td>Cys673</td>
<td>-9.2</td>
<td>-1.2</td>
<td>-9.2</td>
<td>-1.2</td>
<td>-7.7</td>
<td>-0.57</td>
</tr>
<tr>
<td>Asp677</td>
<td>-8.3</td>
<td>-8.0</td>
<td>-8.3</td>
<td>-8.0</td>
<td>-5.2</td>
<td>-3.4</td>
</tr>
<tr>
<td>Arg791</td>
<td>-11.7</td>
<td>-4.7</td>
<td>-11.1</td>
<td>-4.4</td>
<td>-10.3</td>
<td>-3.9</td>
</tr>
<tr>
<td>Arg796</td>
<td>-8.1</td>
<td>-10.1</td>
<td>-8.0</td>
<td>-9.9</td>
<td>-8.0</td>
<td>-9.8</td>
</tr>
<tr>
<td>Arg815</td>
<td>-11.1</td>
<td>-6.5</td>
<td>-11.6</td>
<td>-6.3</td>
<td>-10.2</td>
<td>-5.3</td>
</tr>
<tr>
<td>Thr823</td>
<td>-11.8</td>
<td>-10.5</td>
<td>11.0</td>
<td>-10.0</td>
<td>-9.9</td>
<td>-9.0</td>
</tr>
</tbody>
</table>

* The interaction energy terms $\Delta E_{vdW}$ and $\Delta E_{EL}$ represent the steric and electrostatic contributions to binding, respectively, although here they refer to the contribution to ADP binding that is brought about by each single amino acid residue that is listed. Both $\Delta E_{vdW}$ and $\Delta E_{EL}$ are expressed in kcal mol$^{-1}$.

Results reported in Table 3.3 are consistent with the situation reported by Figure 3.4 – The isoleucine in T670I mutated complex does not affect the stability of the binding.
In conclusion our results suggest that different amino acid substitutions at position 670 of KIT are have different effects toward both ADP and imatinib binding. In particular, the main conclusions of our combined in vitro and in silico efforts can be summarized as follows. First, the substitutions with alanine, serine, and lysine do not introduce strong structural perturbations to either the active or the inactive form of the receptor. Accordingly, these mutated KITs can undergo auto-phosphorylation, but, at the same time, imatinib can bind and inhibit auto-phosphorylation, albeit slightly less efficiently than for wild-type KIT. Second, substitutions with proline and arginine result in a loss of function because of a highly distorted conformation at the ADP-binding site renders the receptor poorly effective in binding ADP and, thus, less prone to auto-phosphorylation. Third, the substitution with Ile results in gain of function because this is the only substitution able to leave the KIT receptor strongly auto-phosphorylated and to make it
imatinib insensitive. In summary, these data explain why, among GIST patients who develop imatinib resistance, the mutation that causes the Thr670Ile substitution is the only mutation that is naturally selected at codon 670. Indeed, isoleucine is the only amino acid substitution that can arise through point mutation at that position that is able to yield a selective advantage to tumors under conditions of imatinib treatment, being T670R and T670P loss of function mutations and T670A, T670R and T670S are not selected for high concentrations of imatinib.

3.3. Study of two new mutations in PDGFRA

This study was conducted in collaboration with the Istituto Tumori of Milan and concerned the characterization of two new mutations in PDGFRA in patients affected by GIST.

Gastrointestinal stromal tumors (GISTs) are the most common mesenchymal tumors of the gastrointestinal tract. Nearly 90% of GISTs are marked by gain-of-function mutations in KIT tyrosine kinase, most frequently located at exons 11 and 9 and, to a minor extent, at exons 13 or 17. In the absence of KIT mutations, GISTs can present mutations of the PDGFRA gene \(^{77-80}\) that are mainly located at exons 12, 18, and 14.

Recent studies have shown that the type and nature of KIT or PDGFRA mutations correlate with the clinical response to Imatinib. GISTs with the most common KIT exon 11 mutations show the highest response rates, whereas the responsiveness of those with KIT mutated at exon 9 is sensitive to the drug dose. In any case, response to imatinib closely correlates with the presence and type of KIT mutation. GISTs lacking mutations in ckit/PDGFRA show much lower response rates to imatinib, if no response at all. For the most common PDGFRA mutation in GISTs, there is evidence of insensitivity to imatinib.

The two mutations presented here are the V561D and the deletion Δ842-845. This study proposes original technical issues because of the lack of a 3D model for the PDGFRA protein (that has been obtained by homology techniques) and because
descriptions of modelling approaches to simulate and characterize aminoacidic deletions are practically absent in literature.

The outcomes of this work have been collected in a paper that is currently submitted: P. Dileo, S. Prici, E. Tamborini, T. Negri, S. Stacchiotti, G.M. Pavan, A. Gronchi, P. Coco, E. Fumagalli, P.G. Casali, S. Pilotti. **Efficacy of Imatinib Mesylate in two Patients with KIT-negative Metastatic Gastrointestinal Stromal Tumors harbouring PDGFRA mutations from exon 18 and 12.** *Mol. Cancer Ther.* 2010. Submitted.

### 3.3.1. PDGFRA model, point mutation and amino acidic deletion

The 3D model structure of the PDGFRA, presently not available in the Protein Data Bank (PDB) structure repository, was built following a validate homology procedure\(^ {81}\) and refined using several energy minimization rounds, using the all-atom force field (FF) parameter sets *parm99*.\(^ {46}\) The imatinib molecule was then docked into the wild-type protein-binding site using Autodock 3.0\(^ {82}\) and the crystal structure of the KIT/imatinib complex as a template (code 1T46.pdb).\(^ {83}\) The resulting complex was further energy minimized to convergence. Eventual missing force field parameters for the inhibitor were taken from our previous work.\(^ {81,84}\)

![Figure 3.5. PDGFRA structure in complex with imatinib. The PDGFRA protein is represented as a shadowed black solid ribbon as well as the imatinib drug. The 561 residue is colored in blue in the WT complex (Val) and in red in the mutated one (Asp).](image-url)
The point mutation was introduced into the wild-type structure of the corresponding PDGFRA/imatinib complex following a well-validated procedure.\textsuperscript{81,84,85}

As already underlined, the Δ842-845 deletion interests a larger zone of the protein. The induced distortion is more consistent than the one induced by a single point mutation as represented by Figure 3.6.

Figure 3.6. WT (a) and Δ842-845 (b) mutated PDGFRA. The amino acidic portion that is deleted is colored in red as the 841 and 846 residues (represented in stick). The PDGFRA protein is represented as a shadowed black solid ribbon.

The Δ842-845 deletion was induced in the binding site deleting the residues form 842 to 845 in the original PDGFRA structure and recreating the peptide bond between residues 841 and 846. The final structure was final minimized and submitted to a longer MD equilibration stage in order to relax the deleted zone and the binding site structure.

### 3.3.2. Results and discussion

The free energies of binding ($\Delta G_{\text{bind}}$) were calculated using the MM-PBSA approach\textsuperscript{50} Table 3.4 reports affinity energies also for T674I mutation in PDGFRA. As already discussed in the previous sections, T674I in PDGFRA is homologue of T670I in KIT and of T315I in Bcr-Abl that was already demonstrated to provoke strong resistance
phenomena toward imatinib. This mutation is not directly involved in the study presented in this section, but is here reported for comparison.

Table 3.4. Free energies of binding of PDGFRA-imatinib complexes. Results are reported for the WT and both the mutated complexes. Values for T674I are reported for comparison. All energetic values are expressed in kcal mol$^{-1}$.

<table>
<thead>
<tr>
<th>complex</th>
<th>$\Delta E_{MM}$ kcal mol$^{-1}$</th>
<th>$\Delta G_{PB}$ kcal mol$^{-1}$</th>
<th>$\Delta G_{NP}$ kcal mol$^{-1}$</th>
<th>$-T\Delta S$ kcal mol$^{-1}$</th>
<th>$\Delta G_{bind}$ kcal mol$^{-1}$</th>
<th>$\Delta \Delta G_{bind}$ kcal mol$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>-75.1 ±1.4</td>
<td>+54.5 ±1.5</td>
<td>-5.9 ±0.2</td>
<td>+17.5 ±0.6</td>
<td>-9.0 ±0.5</td>
<td>-</td>
</tr>
<tr>
<td>T674I</td>
<td>-74.6 ±1.1</td>
<td>+58.1 ±1.2</td>
<td>-5.8 ±0.2</td>
<td>+17.2 ±0.5</td>
<td>-5.1 ±0.7</td>
<td>-3.9</td>
</tr>
<tr>
<td>V561D</td>
<td>-78.5 ±1.2</td>
<td>+56.5 ±1.4</td>
<td>-5.7 ±0.1</td>
<td>+17.7 ±0.9</td>
<td>-10.0 ±0.8</td>
<td>+1.0</td>
</tr>
<tr>
<td>$\Delta 842-845$</td>
<td>-77.2 ±1.3</td>
<td>+51.7 ±1.6</td>
<td>-5.8 ±0.2</td>
<td>+18.5 ±1.1</td>
<td>-12.8 ±1.0</td>
<td>+3.8</td>
</tr>
</tbody>
</table>

Table 3.4 shows that neither the exon 18 deletion mutation DIMH842-845 nor the exon 12 missense mutation V561D substantially affect the affinity of the kinase for imatinib. The size-altering mutation $\Delta 842-845$ is located in the activation loop of the second kinase domain of the receptor. Although involving the deletion of D842, a critical residue known to be involved in imatinib resistance, the exclusion of D846 from this in-frame deletion has two major effects. It results in a conformational readjustment of the activation loop which does not interfere with imatinib binding (Fig. 3.7c). Furthermore, it leaves this residual aspartic at codon 846 to play the role of controlling the swinging movement of the activation loop that is, in turn, associated with a conformational shift of the adenosine triphosphate (ATP) binding pocket from the active conformation to the inactive one, to which imatinib binds selectively. In the case of the missense substitution V561D, although a small, hydrophobic residue is exchanged by a larger, negatively charged one, the position of the substitution and the relevant environment is such that this change is well tolerated in the juxtamembrane domain of the receptor (Fig. 3.7b). This, in turn, does not result in significant alterations of the inactive (or “closed”) conformation of the kinase and, hence, of the imatinib binding site.
These evidences expressed by Fig. 3.7 are clearly supported by the estimated free energy of binding ($\Delta G_{\text{bind}}$) of imatinib to wild-type (Fig. 3.7a), V561D (Fig. 3.7b) and $\Delta$842-845 (Fig. 3.7c) mutated receptors reported in Table 3.4. As we can see from this Table, both mutated receptors are supposed to have higher affinity for imatinib than the wild-type kinase. In fact, the free energy of binding difference $\Delta \Delta G_{\text{bind}}$ between wild-type and each PDGFRA mutant and imatinib, defined as:

\[ \Delta \Delta G_{\text{bind}} = \Delta G_{\text{bind}} \text{(wild-type)} - \Delta G_{\text{bind}} \text{(mutant)} \]

are both positive, indicating a favorable substitution or deletion at the considered positions. In particular, the $\Delta$842-845 mutant kinase isoform is predicted to be substantially more sensitive to imatinib with respect to the wild-type counterpart, in line with the enhanced clinical response to therapy of the patient carrying this mutation. This is an interesting result and is due to the favorable interplay between $\Delta E_{\text{MM}}$ and $\Delta G_{\text{PB}}$ that allows higher affinity for the imatinib drug despite a higher unfavorable entropic term. The $-T\Delta S$ term for $\Delta$842-845 is in average 1 kcal mol$^{-1}$ less favorable than in the other complexes – this is due to the higher constrain induced by the deletion of four amino acids and to the rigidity induced in the binding site.

Molecular modelling was not provided previously for these mutations. Our simulation study showed that the in-frame deletion $\Delta$842-845, and the missense substitution V561D do not result in decreased imatinib sensitivity. In fact, from a structural standpoint, we observed that neither of the two mutations interfere significantly with the conformation of the imatinib binding site; accordingly, the calculated affinities ($\Delta G_{\text{bind}}$) of these mutant isoforms for the inhibitor are higher than that of the wild-type receptor.

Regarding the PDGFRA $\Delta$842-845 mutation, it is interesting to note that it is homologous to the D816V mutation of KIT. The latter is associated with the ATP binding pocket “open” conformation, which prevents response to imatinib. It appears thus evident that not all homologous substitutions should be considered equal.

In conclusion, we provided in vivo evidence that two mutations at exon 18 (involving deletion of residues $\Delta$H842-845) and 12 (V561D) of PDGFRA are sensitive to imatinib,
and demonstrated that molecular modelling results were able to yield a rationale for these clinical findings.

Indeed, these results demonstrated that modelling investigations may constitute a powerful tool to predict imatinib sensitivity, and can support the clinical treatment of GIST mutations not yet functionally investigated.

Figure 3.7  Detail of the binding site of WT (a – green imatinib surface), V561D (b – pink imatinib surface) and Δ842-845 (c – yellow imatinib surface) PDGFRA in complex with imatinib.
3.4. Synergetic effect between primary and secondary mutations in KIT

Gastrointestinal stromal tumor (GIST) represents the majority of soft tissue sarcoma arising from the gastrointestinal tract. Most GISTs encode activating mutations in the kit gene, an important genetic event in the genesis of tumoral pathologies.

Imatinib mesylate (imatinib, Gleevec; Novartis, Basel, Switzerland) has revolutionized the management of patients with advanced disease by targeting the aberrant kinase activity of Kit.

Mutation status of kit in GIST has emerged as a prognostic and predictive factor for imatinib response. At diagnosis, most KIT mutations are within exon 11 or less commonly within exon 9, 13, or 17 (Chen et al., 2004). Regarding exon 13, early reports suggested that tumors harboring these mutations responded suboptimally to imatinib therapy. However, the efficacy of imatinib in this subpopulation has been difficult to analyze due to its rarity with an estimated incidence of approximately 1% (Desai et al., 2007). Previous studies (by the group of Chen) showed that primary mutation of the KIT kinase domain 1 (exon 13, V654A) correlated to imatinib resistance and rapid progression (Heinrich et al., 2006). Desai and colleagues demonstrated that 80% of patients with imatinib-resistant nodules after initial response to imatinib harbored secondary mutations in KIT exons 13 and 17 (Lasota et al., 2008). Thus, KIT exon 13 can be altered by primary or secondary mutation that can confer resistance to imatinib. However, the differential effect on imatinib binding and clinical response for each particular mutation within KIT exon 13 is not fully understood. The subject of this study is a patient with GIST harboring a KIT exon 13 mutation (K642E) who demonstrated a robust and prolonged response to imatinib. This 66-year-old Caucasian man presented to The M. D. Anderson Cancer Center with recurrent GIST measuring approximately 3.5 cm at the gastroesophageal junction and several small intra- and extra-peritoneal implants (Fig. 3.8 A).
3 – Results & Discussion

Figure 3.8. Radiography: (A) Contrast-enhanced CT (computed tomography) of pre-imatinib tumor (arrow) at the gastroesophageal junction. (B) Contrast-enhanced CT of the same tumor after 1 year of imatinib therapy demonstrating decreased tumor size. (C) Contrast-enhanced CT at the time of multi-focal progression of disease.

Pathologic evaluation revealed a highly cellular, spindle cell tumor with brisk mitotic activity. The tumor was diffusely reactive for KIT, CD34, and focally reactive for smooth muscle actin while being negative for desmin, S-100, and pan-cytokeratin. He was treated with 600 mg daily imatinib, and experienced a decrease in tumor burden by 1 year (Fig. 3.8 B). However, after 45 months of imatinib therapy, he had multi-focal progression within the abdominal cavity and a pleural implant (Fig. 3.8 C). He underwent surgical resection to debulk his tumor burden. Characteristic histopathologic features of treatment response to imatinib were not present in these recurrent tumor areas.

In this study we analyzed the influence of primary and secondary mutations on binding stability. In particular the interest was pointed to the synergetic effect that exists between primary and secondary mutations (in this case, K642E and V564A) – together these two amino acidic substitutions create a strong imatinib resistance, higher than the one characterized by the two mutations taken separately.

The results of this study are reported in details in: J.C. McAuliffe, W.L. Wang, G.M. Pavan, S. Pricl, D. Yang, S.S. Chen, A.J.F. Lazar, R.E. Pollock, J.C. Trent. Unlucky

### 3.4.1. Primary and secondary mutation

Mutational analysis of this tumor at the time of imatinib initiation revealed a missense mutation of KIT exon 13, K642E. This mutation is the most common codon affected in KIT exon 13 (Desai et al., 2007). Upon resistance and progression of disease, mutational analysis revealed a secondary KIT exon 13 mutation (V654A) along with the primary K642E mutation. This finding corroborates those seen by Chen and others regarding the V654A mutation and imatinib resistance (Singer et al., 2002).

Thus, we created the molecular models for KIT wild type, in presence of the two mutations alone and in presence of both the point mutations, in order to explore the synergy effect between K642E and V654A.

Figure 3.9 shows the KIT complex in presence of both the point mutations. As described in the Materials & Methods section, the mutations were introduced using the *biopolymer* module of the Insight II suite of programs. After the introduction of the mutations the KIT-imatinib complexes were solvated in a periodic box extending 12 angstroms form the solute. A suitable number of Na\(^+\) and Cl\(^-\) ions were introduced in the system – replacing water molecules where superposition occurred – in order to reproduce the system neutrality and the relevant salt concentration of 150 mM (organic concentration). Four molecular systems were thus created: KIT in complex with imatinib in wild type condition (WT), mutated K642E, mutated V654A and mutated in both positions (Figure 3.9).

### 3.4.2. Results and discussion

After minimization, the molecular systems were equilibrated by running a 10 nanoseconds molecular dynamic (MD) in NPT conditions. The snapshots to perform the binding energy calculations were taken from the last 2 ns of the equilibrated MD trajectories.
Figure 3.9. Depiction of KIT-imatinib complex mutated in position 642E and 654A. The receptor is represented as a blue ribbon. The juxtamembrane region is colored in green and the α-helix in yellow. The K642E and V654A mutation are represented as purple and orange stick respectively. The imatinib drug is represented in ball and stick colored by atoms – its surface is shown as a grey shadow. Water and counterions are not represented for clarity.
The $\Delta G_{\text{bind}}$, and its components, were calculated with the MM-PBSA approach. Table 3.5 reports the thermodynamic binding data for the four molecular complexes expressed in kcal mol$^{-1}$. As already pointed out, the binding energy $\Delta G_{\text{bind}}$ is defined as follows:

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S$$

(3.1)

where $-T\Delta S$ is the entropic term of binding and the enthalpy $\Delta H$ can be defined as:

$$\Delta H = \Delta E_{\text{MM}} + \Delta G_{\text{sol}}$$

(3.2)

In Eq. (3.2) the enthalpy is defined as a sum of the in vacuum interaction energies

$$\Delta E_{\text{MM}} = \Delta E_{\text{EL}} + \Delta E_{\text{solW}}$$

(3.3)

corrected with a solvation term:

$$\Delta G_{\text{sol}} = \Delta G_{\text{PB}} + \Delta G_{\text{NP}}$$

(3.4)

that takes into account the polar and non-polar solvation contribution.

Table 3.5. Free energies of binding of KIT-imatinib complexes. Results are reported for the WT, both the single-mutated complexes and the KIT in presence of primary and secondary mutation. All energetic values are expressed in kcal mol$^{-1}$.

<table>
<thead>
<tr>
<th>complex</th>
<th>$\Delta E_{\text{MM}}$ kcal mol$^{-1}$</th>
<th>$\Delta G_{\text{PB}}$ kcal mol$^{-1}$</th>
<th>$\Delta G_{\text{NP}}$ kcal mol$^{-1}$</th>
<th>$-T\Delta S$ kcal mol$^{-1}$</th>
<th>$\Delta G_{\text{bind}}$ kcal mol$^{-1}$</th>
<th>$\Delta G_{\text{bind}}$ kcal mol$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>-122.9±3.6</td>
<td>+104.4±13.7</td>
<td>-8.8±0.2</td>
<td>+17.0±3.6</td>
<td>-10.2</td>
<td>-</td>
</tr>
<tr>
<td>K642E</td>
<td>-125.9±5.4</td>
<td>+107.5±11.3</td>
<td>-8.7±0.1</td>
<td>+17.1±3.1</td>
<td>-10.0</td>
<td>-0.2</td>
</tr>
<tr>
<td>V654A</td>
<td>-115.2±6.3</td>
<td>+97.6±8.4</td>
<td>-8.8±0.2</td>
<td>+17.7±4.0</td>
<td>-8.7</td>
<td>-1.5</td>
</tr>
<tr>
<td>both</td>
<td>-115.8±3.8</td>
<td>+99.2±8.4</td>
<td>-8.9±0.2</td>
<td>+17.5±3.9</td>
<td>-8.0</td>
<td>-2.2</td>
</tr>
</tbody>
</table>
By comparing the $\Delta G_{\text{bind}}$ of each mutated complex with the one of the WT complex, it is possible to evaluate the direct effect of mutations on binding affinity. Eq. (3.5) defines $\Delta \Delta G_{\text{bind}}$ as the difference between binding energy for the WT complex and the mutated one:

$$\Delta \Delta G_{\text{bind}} = \Delta G_{\text{bind}}(WT) - \Delta G_{\text{bind}}(\text{mutated})$$ (3.5)

In particular, the more negative the $\Delta \Delta G_{\text{bind}}$, the most the mutation destabilize the binding between drug and protein.

![Figure 3.10. Interior view of the KIT binding site. Imatinib is represented with a green surface and the secondary structure of KIT as a dark shadowed ribbon. The residues 654V and 654A are colored in red and blue and are characteristics of the WT and the V564A mutated complexes.](image-url)
Importantly, molecular modelling of the K642E mutant predicts that the negatively charged glutamic acid (in contrast to the positively charged lysine) results in altered interactions of the control c-helix with the KIT juxtamembrane domain which, in turn, leads to an uncontrolled phosphorylation of the kinase, without effecting binding of imatinib. However, the 654 position of KIT lies within the ATP binding pocket and directly participates in imatinib binding (Figure 3.10). Thus, substitution of a valine for a smaller alanine in the 654 position leads to a “void” in the binding site. This gives rise to a global re-arrangement of the binding site that results in a less efficient drug binding.

The calculated free energy of binding ($\Delta G_{\text{bind}}$) between wild-type KIT and imatinib is -10.22 kcal mol$^{-1}$ (Tamborini et al., 2006). The $\Delta G_{\text{bind}}$ for the K642E mutant is marginally less at -10.03 kcal/mol. The $\Delta G_{\text{bind}}$ for the KIT V654A mutant is -8.7 kcal mol$^{-1}$ – about one order of magnitude decrease in drug affinity, being the relationship between $\Delta G_{\text{bind}}$ and affinity IC50 expressed by Eq. (3.6).

$$
\Delta G_{\text{bind}} = -RT \ln(k) \approx -RT \ln(IC50)
$$

being IC50 defined as:

$$
IC50 = k_i \left(1 + \frac{s}{k_m}\right)
$$

where $k_i$ is the binding affinity of the inhibitor, IC50 is the functional strength of the inhibitor, $[s]$ is the substrate concentration and $k_m$ is the affinity of the substrate for the enzyme.

The half maximal inhibitory concentration (IC50) is a measure of the effectiveness of a compound in inhibiting biological or biochemical function. Often, the compound in question is a drug candidate. This quantitative measure indicates how much of an inhibitor is needed to inhibit a given biological process (i.e. the activity of KIT receptor) by half. In other words, it is the half maximal (50%) inhibitory concentration (IC) of a substance (50% IC, or IC50).
Given the logarithmic relation that exists between IC50 and $\Delta G_{\text{bind}}$, each $\sim1.4$ kcal mol$^{-1}$ of difference in the $\Delta G_{\text{bind}}$ there is an order of magnitude in the IC50 index, and thus in the affinity of the protein for the drug.

The $\Delta G_{\text{bind}}$ for the KIT K642E/V654A double mutant is -8.04 kcal mol$^{-1}$ indicating a 1–2 order of decrease in affinity for imatinib. Thus, the two mutations work together to enhance the resistance to imatinib as compared to each mutation alone.

In conclusion, this GIST patient harboring KIT exon 13 mutation K642E benefited from imatinib. Treatment selected for the double mutation that conferred resistance and eventuated death. Despite the minimal 11 codons separating these loci, the resultant phenotype is imatinib-sensitive for K642E and imatinib-resistant for K642E/V654A. As GIST KIT mutation reporting becomes more prevalent, documentation of the specific mutation is important to distinguish those patients with probable imatinib resistance from those that may benefit from imatinib for many years.

3.5. Characterization of L576P mutation in the KIT-imatinib complex

The interest in the characterization and the study of the L576P mutation was born in collaboration with the M.D. Anderson hospital of Houston and Istituto Tumori of Milan almost at the same time.

The group of Prof. Trent at Houston reported a case of a patient affected by metastatic melanoma. The patient was not responding to the treatment with imatinib, due to the presence of a very rare mutation: the L576P. As extreme measure, they treated the patient with dasatinib – a drug commonly used to inhibit the BCR-ABL protein in patients affected by Chronic myeloid leukemia (LMC). The patient responded surprisingly well after only three week of treatment.

This case is intrinsically interesting because mutations arising in the juxtamembrane tract are well known in literature to activate the tyrosine kinase domain of these proteins but still to respond to the inhibitor action of imatinib. This does not happen for this mutation L576P, that causes a strong resistance effect toward the activity of this drug.
That is why the first study presented in this thesis in this framework is a collaboration with the Istituto Tumori of Milan concerning a characterization of L576P mutation in KIT complexated with imatinib.


### 3.5.1. Mutations in the juxtamembrane domain

L576P is a rare KIT mutation often reported in cancers other than gastrointestinal stromal tumors (GIST). In GISTs, it correlates with features linked to an aggressive phenotype, eventually resulting in secondary mutations. In vitro analyses point out that L576P KIT is constitutively activated, and shows poor imatinib sensitivity. In the article mentioned above, histological, immunohistochemical, and biochemical analyses, coupled with mutational-molecular analysis and fluorescence in situ hybridization, were applied to surgical specimens. In parallel, the affinities of wild-type, L576P KIT, and Δ559/KIT for imatinib were estimated by *in silico* studies. Despite imatinib treatment and the apparent clinical-imaging response, the detected histological response was very low. Computer modelling proved that L576P KIT is two times less sensitive than the wild-type counterpart and considerably less affine to imatinib than the sensitive Δ559/KIT. Accordingly, the modelling evidence strongly supports the lack of tumoral regression we observed at the histological level.

The molecular profile of most gastrointestinal stromal tumors (GIST) is characterized by the mutually exclusive oncogenic mutation of two members of the type III tyrosine kinase receptors: KIT and PDGFRA. This finding provided the rationale for tyrosine kinase inhibitor (TKI) treatments in these tumors, which are highly resistant to conventional therapies. KIT mutations are the most frequently encountered activating mutations in GISTs, being present in about 80% of these malignancies. They mainly affect exon 11 in the kinase juxtamembrane domain (JMD), are mostly in heterozygous status, and include deletions, insertions, and missense mutations. Importantly, mutations
in the KIT JMD are usually the most responsive to the ATP-competitive small molecule inhibitor imatinib.\textsuperscript{86,87} Among the missense mutations identified in GISTs, Leu576Pro (L576P) is very rare, accounting for 8\% of all exon 11 point mutations, which, in turn, range between 16.3 and 26.5\%.\textsuperscript{87,88} Here we describe two cases of imatinib-treated and surgically resected GIST harboring L576P KIT exon 11 mutation, which, despite clinical and imaging evidence of regression, microscopically showed a low, minimal response rate in most of the tumoral deposits.

Here we proposed a comparison between the KIT WT, the Δ559 mutated KIT and the L576P mutated KIT in complex with imatinib.

As said Δ559 is a mutation well known to activate the KIT but also for its high affinity for imatinib drug. Many other mutations that interest the juxtamembrane tract are reported to have similar effect (V559D) – as soon as the KIT express these kind of mutations in the amino acidic chain, the patient can be successfully treated with imatinib. L576P however show an unexpected resistance for imatinib, differently from the other mutations that interest the same region of the protein structure.

Molecular modelling supported experiments highlighting the origins of the low affinity of imatinib for KIT mutated L576P.

### 3.5.2. History of the cases studied

**Mutational Clinical History: Case 1**

A 60-year-old male presented with a peritoneal metastasis from a high risk ileal GIST resected 1 year before. He was started on imatinib 400 mg/day, resulting in a complete remission. One year after, he developed a focal peritoneal progression and was operated on. The progressing lesion was resected, and, unexpectedly, many peritoneal implants (not visible on CT scans) were found, most of which were removed. He was further continued on imatinib 800 mg/day. Sixteen months after surgery he developed a generalized progression, which was also not responsive to nilotinib. He died of abdominal liver and peritoneal metastases 3 years after imatinib onset.
**Clinical History: Case 2**

A 70-year-old male presented with a large duodenal GIST. A preoperative treatment was started in order to avoid a major surgical procedure (pancreatic-duodenectomy). He received imatinib 400 mg/day for 10 months, obtaining a radiological response (reduction in tumor volume and density on CT scan). He was then operated on, and a conservative duodenal resection could be done. A postoperative treatment with imatinib 400 mg/day for one more year was planned.

### 3.5.3. Results

L576P is a rare mutation located in the JMD of KIT described in GISTs, melanomas, and canine GISTs and mastocytomas. We observed three L576P mutations among 332 profiled GISTs, of which we report upon two cases here. Case 1 represents an intrabdominal recurrence of a high risk GIST arising from the ileum, whereas case 2 is a primary post-treatment high risk tumor of the duodenum. The functional significance of L576P substitution and its imatinib sensitivity have been previously investigated on cell lines. In particular, the cells expressing the L576P KIT c-DNA revealed spontaneous KIT phosphorylation, and in vitro sensitivity tests highlighted the need of a tenfold higher dose of imatinib to switch off activated L576P KIT with respect to V599D KIT.

In terms of KIT activation, in case 1 the strong KIT expression-activation is consistent with the presence of secondary mutations that we were ultimately unable to detect, despite extensive searches with the available technique (definitely less sensitive than the recently proposed allele specific PCR). In case 2, the expression-activation of KIT on primary tumor was not particularly high, a finding otherwise consistent with the high heterogeneity of both KIT expression and phosphorylation previously reported in naive and treated GISTs. However, the evidence of post treatment phosphorylation, along with the low rate of histological regression, support the notion that imatinib 400 mg/day was unable to switch off KIT.

In order to find a rationale for these findings, we did “in silico” experiments on two mutant KITs along with wild-type in complex with imatinib. The present MD simulations...
show that: (i) the L576P KIT is considerably less affine to imatinib than the very sensitive Δ559/KIT (or V599D KIT, as experimentally verified by Antonescu and colleagues; ref. 11), and (ii) this KIT isoform overall results in a kinase that is two times less sensitive to imatinib with respect to its wild-type counterpart.

Figure 3.11. KIT WT (a) and superposition of Δ559 KIT and L576P KIT (b). The WT KIT is represented as a cyan ribbon, the Δ559 KIT and L576P mutated KIT are colored in purple and grey respectively. The imatinib drug is represented in stick, water molecules and ions are omitted for clarity.

Figure 3.11 represents the different effect that these two mutations have on the protein structure. The deletion of the Val559 in Δ559 KIT create as a sort of “domino effect” a void in the deeper part of the binding site. The picture shows clearly how the drug is able to better accommodate within the binding pocket, due to a rearrangement of the whole juxtamembrane domain caused by the Δ559 deletion. On the other hand, L576P KIT (in gray in the picture), characterizes a completely different scenario. The Pro576 (in stick in the picture), due to the higher steric occupancy and lower flexibility respect to the
original Leu576 of the WTW complex, pushes the α-helix toward the binding site, reducing the space at disposal for the imatinib and creating a worse thermodynamic condition – van der Waals and electrostatic repulsions – for the imatinib binding event.

Table 3.6. Free energies of binding of KIT-imatinib complexes. Results are reported for the WT and L576P and Δ559 mutated KIT. All energetic values are expressed in kcal mol⁻¹.

<table>
<thead>
<tr>
<th>complex</th>
<th>ΔE_{EL} (kcal mol⁻¹)</th>
<th>ΔE_{vdW} (kcal mol⁻¹)</th>
<th>ΔG_{PB} (kcal mol⁻¹)</th>
<th>ΔG_{NP} (kcal mol⁻¹)</th>
<th>-TΔS (kcal mol⁻¹)</th>
<th>ΔG_{bind} (kcal mol⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>-45.2±3.7</td>
<td>+76.0±3.6</td>
<td>+104.4±13.7</td>
<td>-8.7±0.2</td>
<td>+17.5±3.4</td>
<td>-8.1</td>
</tr>
<tr>
<td>L576P</td>
<td>-40.6±6.0</td>
<td>-73.6±3.5</td>
<td>+98.3±7.2</td>
<td>-8.8±0.1</td>
<td>+17.0±3.9</td>
<td>-7.8</td>
</tr>
<tr>
<td>Δ559</td>
<td>-44.4±4.3</td>
<td>-78.2±3.6</td>
<td>+105.0±10.9</td>
<td>-8.7±0.1</td>
<td>+16.9±3.3</td>
<td>-9.4</td>
</tr>
</tbody>
</table>

The application of the MM-PBSA approach to the L576P/KIT in complex with imatinib resulted in a calculated free energy of binding between inhibitor and kinase, ΔG_{bind}, of −7.8 kcal mol⁻¹. For comparison, the estimated affinity of the imatinib sensitive Δ559 isoform of KIT toward the inhibitor was ΔG_{bind} = −9.4 kcal mol⁻¹, whereas the corresponding value for the wild-type KIT was −8.1 kcal mol⁻¹. Figure 3.11 shows the equilibrated molecular dynamics (MD) models of the wild-type and L576P KIT in complex with imatinib, respectively. Structurally, in the L576P KIT the induced fit for imatinib binding is no longer preserved, because of a conformational change that resembles the more active form of KIT for which imatinib is known to have lower binding affinity.

Moreover, data of Table 3.6 show that the resistive phenomenon in L576P mutated case is strongly conditioned by the non bond interaction energy – the pure mechanical interaction. This is the result of the modification in the binding site structure caused by the L576P mutation. As said, in fact, Pro576 pushes the α-helix toward the imatinib and thus create a less favorable energetic situation for the drug within the binding site of KIT.

In details, our calculations confirm that the Δ559 mutant KIT is more than one order of magnitude more affine to imatinib than the naive kinase. Indeed, when the kinase is in the
closed inactive form, the only form that imatinib is able to bind to, the slight distortion induced by one residue deletion in the JMD conformation causes a conformational rearrangement that also involves the ATP-binding pocket, allowing for a larger space and, hence, a better inhibitor fitting into its binding site. Thus, the modelling evidence gives strong support to the lack of tumoral regression that we observed at a histological level. On the basis of this study, the structural similarity of nilotinib with imatinib, and the analogous binding mode to KIT of these three small molecule ATP-competitive inhibitors (data are not shown), also account for the failure of their action on L576P mutated KIT.

3.6. Inhibition of KIT L576P: imatinib vs. dasatinib

In the previous section, the L576P mutation in KIT complex with imatinib was extensively studied. In this section the interaction of KIT with other inhibitor will be explored.

As said, the study already discussed not the first one faced on L576P in a time scale, however we presented it as a first case studied in order to give more clarity to the other outcomes presented further.

Together with the collaboration with Istituto Tumori of Milan in this framework, we conducted parallel studies with the M.D. Anderson hospital of Houston, Texas. This activity was focused on a different case – a patient affected by melanoma, expressing a L576P mutated KIT, that was not responding at all to the imatinib therapy. The hospital tried to treat this case with another inhibitor drug (dasatinib) that is usually adopted for the treatment of chronic myeloid leukemia (LMC). The patient responded surprisingly well, showing a consistent remission of the tumoral activity after only three weeks of treatment.

In this framework, molecular modelling techniques were used to understand the motivation of this imatinib-resistive and dasatinib-respondent phenomenon. As will be described also in further sections, the results of this study have had important further consequences.
3.6.1. Introduction

Point mutations in the KIT receptor tyrosine kinase gene have recently been identified in mucosal, acral lentiginous, and chronically sun-damaged melanomas.

Patients with metastatic melanoma have a median survival of 6 to 8 months. Unfortu- nately multiple clinical trials conducted at M.D. Anderson with chemotherapy, immunotherapy, and biochemotherapy have failed to significantly improve survival. Protein kinase inhibitors are beneficial in diseases with highly prevalent oncogenic events (e.g., CML, GIST). More than 50% of melanomas arising from areas without chronic sun damage harbor activating mutations of BRAF. However, BRAF mutations are extremely rare in other melanoma subtypes: acral lentiginous (AL), chronic sundamaged (CSD), and mucosal. Interestingly, amplification of chromosomal region 4q12 was seen frequently and selectively in AL, CSD, and mucosal melanomas. Interrogation of candidate genes in this region led to the discovery of frequent mutations and or amplifications of the KIT tyrosine kinase receptor gene in these subtypes.
The identification of KIT mutations in melanoma has direct therapeutic implications. Activating KIT mutations are present in about 85% to 90% of gastrointestinal stromal tumors (GIST).\(^\text{102}\) As already pointed out, treatment with the KIT inhibitor imatinib significantly improved survival in GIST patients, and it is now the standard of care for this disease.\(^\text{103}\) Three previous clinical trials using imatinib in unselected melanoma patients failed to show clinical benefit.\(^\text{104-106}\) However, it is possible that KIT inhibitors will be beneficial to the subset of melanoma patients with KIT mutations. Recent case reports have described clinical responses following treatment with imatinib in two patients with metastatic mucosal melanoma with KIT K642E and KIT 577 duplication mutations, respectively.\(^\text{107,108}\) M.D Anderson hospital reported a complete response in a metastatic mucosal melanoma patient with a KIT V560D mutation who was treated with a sorafenib-based regimen, which also inhibits KIT.\(^\text{109}\)

Although these early responses are encouraging, there are characteristics of KIT mutations in melanoma that suggest imatinib resistance may be an issue in the treatment of these patients. In GIST, the majority (~80%) of KIT mutations occur in the juxtamembrane regulatory domain encoded by exon 11. Most of these mutations have been characterized both in vitro and clinically as being imatinib-sensitive. In contrast, imatinib resistant mutations occur in exons 13 and 17, the kinase domains of the protein. These mutations are rare in GIST (exon 13, <1%; exon 17, 1%). The KIT mutations identified in melanomas occur in the same exons as those affected in GIST. However, there is a greater prevalence of mutations in exon 13 (20%) and exon 17 (10%).\(^\text{101,107-115}\) Additionally, in GIST >90% of the observed mutations are deletions or insertions, whereas >90% of the KIT mutations in melanoma are point mutations. The L576P KIT mutation, which is the most common KIT mutation reported to date in melanoma (~30-40% of mutations), is located at the C-terminus of KIT exon 11, although KIT exon 11 deletions in GIST occur mostly at the N-terminus of KIT exon 11.\(^\text{101}\) These differences in the type and localization of KIT mutations may impact drug efficacy.

Here we report the identification and characterization of the first human melanoma cell line with a L576P KIT mutation. The L576P mutation is the most common KIT mutation identified to date in melanoma, and represents approximately 30% to 40% of the reported
Results & Discussion

point mutations. Despite the location of the L576P mutation in exon 11, which is generally associated with imatinib sensitivity in GIST, we observed that the cells were resistant to the growth inhibitory effects of imatinib. In contrast, the cell line was sensitive to another drug — the dasatinib. We report here the results of molecular modelling studies to investigate the structural effects of the mutation on the interaction with KIT inhibitors.

3.6.2. A different functionality: the results

With the same procedure followed in the previous cases, we created the WT and L576P mutated complexes with both imatinib and dasatinib. Since in the protein data bank the pdb structure of KIT-dasatinib complex was not present, but the homologue BCR-ABL-dasatinib was available, we created it by “superposition”. We superimposed the atomic structure of the binding site of the two proteins and “passed” the dasatinib drug from the Abelson protein to the KIT receptor.

After this step, a preliminary minimization, followed by a pre-equilibration MD step was necessary in order to obtain a reliable starting configuration for the systems.

Figure 3.13. Two MD snapshots of WT KIT in complex with imatinib (a) and dasatinib (b). Within the KIT structure, helixes are represented in green, strands in yellow, and the
rest in blue. The drugs are represented in stick and colored by atoms. The Leu576 is represented in red stick.

After this step, Molecular dynamics (MD) simulations were carried out using the AMBER 9 (Sander.MPI and pmemd module) suite of programs and the parm99 all-atom force field working in parallel on 64 processors of the IBM/BCX calculation cluster of the CINECA calculation center of Bologna, as described previously.

Figure 3.13 shows two snapshots of the MD simulations of the WT complexes taken from the equilibrated phase of the dynamic trajectories.

It is worth noting the different positioning of the two drugs inside the binding site respect to the mutation. The two drugs work in a different way in the inhibition – imatinib penetrates deeper in the site while dasatinib stops the external portion of the binding pocket.

Figure 3.14. Representation of the different binding mode of imatinib (a) and dasatinib (b). It is evident how imatinib penetrate deep in the binding site, positioning under the α-helix (in green below), while dasatinib occupies the external zone of the binding pocket.
Figure 3.14 represents schematically the different functionality of imatinib and dasatinib drugs. Conceptually, imatinib, that penetrate deeply in the binding site, is more strongly affected by a mutation – i.e. L576P – that interest the zone of the α-helix (colored in green in the below schemes of Figure 3.14). This finds evidences also in the affinity energy evaluations reported in Table 3.7.

Table 3.7. Free energies of binding of KIT-imatinib and KIT-dasatinib complexes. Results are reported for the WT and L576P mutated KIT. All energetic values are expressed in kcal mol\(^{-1}\).

<table>
<thead>
<tr>
<th>complex</th>
<th>ΔE(_{\text{MM}})</th>
<th>ΔG(_{\text{PB}})</th>
<th>ΔG(_{\text{NP}})</th>
<th>-TΔS</th>
<th>ΔG(_{\text{bind}})</th>
<th>ΔΔG(_{\text{bind}})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kcal mol(^{-1})</td>
<td>kcal mol(^{-1})</td>
<td>kcal mol(^{-1})</td>
<td>kcal mol(^{-1})</td>
<td>kcal mol(^{-1})</td>
<td>kcal mol(^{-1})</td>
</tr>
<tr>
<td><strong>imatinib</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>-122.9±3.6</td>
<td>+104.4±13.7</td>
<td>-8.8±0.2</td>
<td>+17.0±3.6</td>
<td>-10.3</td>
<td></td>
</tr>
<tr>
<td>L576P</td>
<td>-114.2±6.3</td>
<td>+98.3±7.2</td>
<td>-9.0±0.1</td>
<td>+17.2±3.9</td>
<td>-7.8</td>
<td>-2.5</td>
</tr>
<tr>
<td><strong>dasatinib</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>-98.7±6.5</td>
<td>+79.1±8.3</td>
<td>-7.0±0.2</td>
<td>+17.0±3.0</td>
<td>-9.61</td>
<td></td>
</tr>
<tr>
<td>L576P</td>
<td>-96.5±7.8</td>
<td>+76.2±10.7</td>
<td>-7.0±0.2</td>
<td>+17.3±4.1</td>
<td>-9.93</td>
<td>+0.32</td>
</tr>
</tbody>
</table>

Table 3.7 shows how the KIT-imatinib complex is strongly affected by the presence of the L576P mutation – the ΔΔG\(_{\text{bind}}\) value is of -2.5 kcal mol\(^{-1}\), meaning that, following Eq. 3.6, the imatinib is almost two orders of magnitude less affine to the L576P mutated KIT than to the WT one. On the other hand, the KIT-dasatinib complex seems not to feel at all the presence of the mutation in position 576 of the KIT amino acidic chain. This is basically due to the different functionality of the two drugs. Dasatinib stops at the external zone of the binding site – the distance between the drug and the L576P mutation is higher than in the imatinib case.

This is the main reason of the imatinib resistance phenomenon versus the dasatinib sensitivity: The different structural complexation results in a different energetic situation within the binding site that will be described in details in the next section.
3.6.3. Inside the binding site

The conclusion that the different affinity (high for dasatinib and low for imatinib) of the mutated complexes is related to the different penetration of the two drugs is exhaustively represented by Figure 3.15. By superimposing the backbones of the WT and mutated structures of the two different complexes it is immediate to realize that the L576P mutation (in red in the picture) creates a consistent destabilization in the imatinib complex, that is not present in the dasatinib one (purple and green dasatinib molecules are practically superimposed).

In fact even the substitution of a single amino acid (L576P) creates a small reassessment in the structure of the narrow α-helix. Due to the deep conformation of imatinib inside the binding pocket of KIT, the “tail” of the drug is immediately below the α-helix, and its conformation is strongly affected by the presence of the L576P mutation.

Figure 3.15. Backbone superposition of WT-KIT and L576P-KIT in complex with imatinib (colored in blue and orange respectively) and of WT-KIT and L576P-KIT in complex with dasatinib (green and purple respectively). The drugs as well as the 576 residue are represented in stick and the KIT structure is represented as a solid ribbon. Hydrogen atoms and water molecules are omitted for clarity.
In the dasatinib case however this does not happen. Because of the external conformation assumed by the dasatinib at the equilibrium, the α-helix of KIT is exactly between the drug and the point mutation. In this case the α-helix acts as a screen that insulates the binding pocket from the destabilizing action of the L576P mutation, resulting in an invariant affinity for the drug - $\Delta \Delta G_{\text{bind}}$ value is of +0.34 kcal mol$^{-1}$.

This explanation motivates the data of Table 3.7, demonstrating why L576P creates resistance phenomena toward imatinib and does not have any particular effect in the dasatinib complex.

However, this is only a qualitative demonstration. In order to achieve a quantitative insight inside the binding site another step was necessary. By performing a per-residue decomposition of the binding energy of each complex it was possible to highlight the residues that most strongly stabilize the drug in the binding site. Moreover, by comparing the interactions with the drugs for each of these residues in the WT and in the mutated complex in terms of non-bond energies (electrostatic and van der Waals), it was possible to quantify the effect that the L576P mutation has in destabilizing the binding site in terms of single residues drug attraction.

Importantly, with this analysis it was possible to correlate the structural and energetic consequences of the point mutation – a real “bridge” in the structural and energetic study of these drug resistance phenomena.

In fact, being the protein-ligand interaction energy represented as:

$$\Delta E_{\text{MM}} = E_{\text{MM-complex}} - (E_{\text{MM-protein}} + E_{\text{MM-ligand}}) \tag{3.8}$$

it was possible to decompose the total interaction energy on a per-residue base.

$$\Delta E_{\text{MM}} = \sum_{i=1}^{n} \Delta E_{\text{MM}} (i) \tag{3.9}$$

The interaction between each residue of KIT and the drugs was calculated both in the WT ($\Delta E_{\text{MM-WT} (i)}$) and in the mutated case ($\Delta E_{\text{MM-L576P} (i)}$). Thus, it was possible to
calculate the effect the L576P mutation has in destabilizing the action of each residue of the binding site according to Eq. 3.10:

\[
TGAS(i) = \left[ \Delta E_{MM}(i) - WT \right] - \left[ \Delta E_{MM}(i) - L576P \right]
\]

The huger the TGAS value, the more consistent the role played by the mutation in destabilizing (if TGAS < 0 – the favorable interaction \( \Delta E_{MM} \) is a negative number according to Eq. 3.8) or in stabilizing (if TGAS >0) the binding between drug and a precise amino acid of the protein.

Moreover, TGAS is composed as:

\[
TGAS(i) = TVDW(i) + TELE(i)
\]

Data related to the KIT-imatinib complex are reported Figure 3.16a (and in the related Table – Figure 3.17b).

Figure 3.16. a) Superposition of the crucial residues of the binding site of KIT-imatinib complex in WT (blue) and L576P (orange) conditions. The backbone of the KIT receptor is represented as a grey-shadowed ribbon. The imatinib drug is represented in ball and stick and the principal amino acids in stick. Water molecules and ions (as well as hydrogen atoms) are omitted for clarity. b) Interaction energy values (TGAS is defined
by Eq. 3.11) between each of the more important amino acids of the binding site (more important residues in the stabilization of the drug within the binding site of KIT) and the imatinib drug. Energies are expressed in kcal mol$^{-1}$.

Figure 3.16 represents graphically, and express quantitatively (with the support of per-residues interaction energy values reported in panel b) the destabilization within the binding site caused by the L576P mutation. To go through the Table, it is worth to underline that the more negative the TGAS, TVDW and TELE, the higher the destabilization induced by the single mutation in the attraction power of each single residue. Importantly, the substitution of a leucine with a proline in position 576 creates a strong repulsion especially in the residue of the upper zone of the binding site. This is consistent, and gives a quantitative insight into the observations made previously – the mutation induces a small shifting of the $\alpha$-helix that create a movement in the “tail” of the drug. This is caused by a sort of “domino effect” to the repulsion (electrostatically driven –TELE) between the drug and the main residues closer to the $\alpha$-helix (640, 623 and 640 – very negative TGAS values).

It is also interesting to note that (last two rows of Figure 3.16b) the getting worse of the six main residues is also higher than the total one (sum of all the residues of the KIT receptor). It means that the binding site of KIT is definitely strongly destabilized by the L576P mutation, that the rest of the proteine structure attempts to compensate with a general reassessment, but without success.

Figure 3.17 provide the same insight into the KIT binding site but in complex with dasatinib.

The situation induced in the binding site by the dasatinib drug is completely different. As said, this drug stops in the external part of the binding pocket, acting as a “stopper”. Figure 3.17 shows how the principal residues of the KIT binding site are only slightly affected by the L576P mutation (total TGAS related to the seven main amino acids ok -1.47 in kcal mol$^{-1}$). This getting worse is an order of magnitude smaller than the one reported by the imatinib complexes (Figure 3.16b – TGAS = -12.55 in kcal mol$^{-1}$). That is due to the mutual position of dasatinib drug. L576P mutation and the $\alpha$-helix.
Figure 3.17. a) Superposition of the crucial residues of the binding site of KIT-dasatinib complex in WT (blue) and L576P (orange) conditions. The backbone of the KIT receptor is represented as a grey-shadowed ribbon. The dasatinib drug is represented in ball and stick and the principal amino acids in stick. Water molecules and ions (as well as hydrogen atoms) are omitted for clarity. b) Interaction energy values (TGAS is defined by Eq. 3.11) between each of the more important amino acids of the binding site (more important residues in the stabilization of the drug within the binding site of KIT) and the dasatinib drug. Energies are expressed in kcal mol^-1.

In the dasatinib complexes, in fact, the α-helix stays exactly between mutation and drug, acting as a screen. The drug hardly feels the presence and the effect of the L576P mutation.

From the clinical point of view, this analysis can give an interpretation of the radical different response of the patient to the treatment with these two different drugs.

While imatinib demonstrated to be quite ineffective in the treatment of the patient, dasatinib provokes an almost complete remission of the tumoral activity (Figure 3.17) after only three weeks of therapy.

Concluding, this analysis provides a representative, qualitative and quantitative interpretation of the effect induced in a single mutation (L576P). This gives a complete interpretation of the imatinib resistance and dasatinib responsive effects in the characterization of the new point mutation L576P in KIT.
Figure 3.18. PET/CT images of the pelvis zone of the patient. a) Fluorescences: FDG-avid right pelvic melanoma tumors harboring the L576P KIT mutation before dasatinib treatment (patient under imatinib treatment). b) After three weeks on dasatinib treatment.
3.7. KIT L576P: the functionality of different inhibitors

The study already described in the previous sections raised immediately the necessity to explore and analyze the effect of other inhibitor drugs in the treatment of the new mutation L576P.

The M.D. Anderson hospital suggested other three drugs that are extremely important in the treatment of tumoral diseases: nilotinib, sorafenib and dasatinib. The aim of this study was to find a rationale in the functionality of different drugs.

Figure 3.19. The different structure of nilotinib, sorafenib and sunitinib.

Figure 3.19 reports the molecular structures of these three drugs – common and widely used inhibitors of the several kind of tumoral diseases. Nilotinib is a drug commonly used for the treatment of myeloid leukemia (CML) in case of imatinib resistance. Sorafenib and sunitinib are usually employed in the treatment of renal cell carcinoma and imatinib-resistant gastrointestinal stromal tumors (GISTs).

Figure 3.20 summaries the aim of this study – the comparison between these three drugs with dasatinib and imatinib, and of their effectiveness in successfully bind and inhibit the KIT receptor in presence of L576P mutation.
Figure 3.20. Representation of the aim of this study – to compare the efficiency of the different inhibitors with the imatinib drug in binding the KIT receptor.

The results of this study, in collaboration with the M. D. Anderson hospital of Houston, have been collected and analyzed. A manuscript related to this matter is presently under preparation.

3.7.1. Two different penetrations

Connecting to the study exposed in the last sections, the first, obvious step that has been done in this framework was to evaluate the different functionalities of these new drugs. Figure 3.21 represent the drugs in terms of imatinib-like or dasatinib like inhibitors, based on the different penetration of these molecules within the KIT binding site.

By comparing the equilibrium configuration assumed in the binding site of KIT, it is immediate the division of these drugs in the two groups. It is worth noting how the van der Waals surfaces of drugs belonging to the same family are almost perfectly coincident.
In fact, sorafenib and nilotinib (in yellow and blue respectively in Figure 3.21a) are positioned very close to the L576P mutation, and immediately underneath the α-helix of KIT – similarly to imatinib. On the other hand, sunitinib assumes a more external position (cyan – Figure 3.21b) and acts as a “stopper” – similarly to dasatinib – toward the binding site.

Figure 3.21. Inhibitors can be divided in two different groups – imatinib-like (a) and dasatinib-like (b) drugs due to the different functional penetrations inside the binding pocket of the KIT receptor. In the picture, imatinib, sorafenib and nilotinib (a) are superimposed and represented in green, yellow and blue respectively, dasatinib and sunitinib (b) are colored in pink and cyan respectively. The KIT structure is represented as a gray-shadowed ribbon and the L576 is colored in red for the KIT-imatinib-like (a) and in green for the KIT-dasatinib-like complexes. Water molecules and counterions, as well as hydrogen atoms, are omitted for clarity.

3.7.2. Outcomes

With this in mind, it possible to guess that drugs that belong to the same “family” should behave similarly in the inhibition of the KIT kinase. The molecular systems were created following the same procedure previously illustrated as well as the simulation
runs. In order to have a measurement of the drug-kinase affinity for the WT and the L576P mutated complexes, the free energies of binding were calculated. Table 3.8 reports the $\Delta G_{\text{bind}}$, as composed by an enthalpic ($\Delta H$) and an entropic term ($-T\Delta S$) – all energies are expressed in kcal mol$^{-1}$.

Table 3.8. Free energies of binding of KIT in complex with various inhibitors (1st column). Results are reported for the WT (a) and L576P (b). All energetic values are expressed in kcal mol$^{-1}$.

<table>
<thead>
<tr>
<th>a) KIT-WT inhibitor</th>
<th>$\Delta H$ kcal mol$^{-1}$</th>
<th>$-T\Delta S$ kcal mol$^{-1}$</th>
<th>$\Delta G_{\text{bind}}$ kcal mol$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>imatinib</td>
<td>-25.6±4.5</td>
<td>+17.5±3.4</td>
<td>-8.1</td>
</tr>
<tr>
<td>nilotinib</td>
<td>-25.8±4.1</td>
<td>+17.4±3.8</td>
<td>-8.3</td>
</tr>
<tr>
<td>sorafenib</td>
<td>-25.7±3.4</td>
<td>+17.6±3.1</td>
<td>-8.0</td>
</tr>
<tr>
<td>dasatinib</td>
<td>-26.6±3.8</td>
<td>+17.0±3.0</td>
<td>-9.6</td>
</tr>
<tr>
<td>sunitinib</td>
<td>-27.4±4.0</td>
<td>+17.6±3.5</td>
<td>-9.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b) KIT-L576P inhibitor</th>
<th>$\Delta H$ kcal mol$^{-1}$</th>
<th>$-T\Delta S$ kcal mol$^{-1}$</th>
<th>$\Delta G_{\text{bind}}$ kcal mol$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>imatinib</td>
<td>-24.8±3.3</td>
<td>+17.0±3.9</td>
<td>-7.8</td>
</tr>
<tr>
<td>nilotinib</td>
<td>-24.5±3.1</td>
<td>+17.3±3.3</td>
<td>-7.2</td>
</tr>
<tr>
<td>sorafenib</td>
<td>-24.4±3.3</td>
<td>+17.3±3.5</td>
<td>-7.1</td>
</tr>
<tr>
<td>dasatinib</td>
<td>-27.2±3.8</td>
<td>+17.3±3.4</td>
<td>-9.9</td>
</tr>
<tr>
<td>sunitinib</td>
<td>-27.2±3.8</td>
<td>+17.7±3.9</td>
<td>-9.6</td>
</tr>
</tbody>
</table>

Table 3.8 reports the affinity energies for both the KIT-imatini-like (a) and the KIT-dasatinib-like complexes (b). Data are consistent with the initial guess – drugs belonging to the same “family” show similar affinity toward the same kinase (KIT).

This conclusion is extremely interesting. The possibility to, also approximately, guess the efficiency of a certain drug starting from the mutual positioning toward a certain mutation opens new frontier in the program of ad hoc therapies based on the mutation
situation of the single patient. This is a very promising starting point, but of course, since even a single mutation can induce there “domino effects” destabilizing the whole binding site, the prevision of the functionality of a certain drug in this cases remains a non-trivial issue. In this framework, a further step is presently on going, and will be described in details in the next sections.

3.8. KIT mutations: benchmark of affinity toward imatinib

As previously said, even if the functionality of a certain drug (i.e. penetration in the binding site) in the inhibition of a kinase is well known, the prevision of the effect caused by the rising of even a single amino acidic substitution in the protein chain is not easy to predict. As seen in all the cases described in previous sections, a substitution of a precise amino acid with another one provokes a general reassessment of the entire binding site due to non-bond repulsions, or attraction caused by the different shape of the amino acid side chains – the so-called “domino effect”. It is thus necessary to create a benchmark of the most frequent mutations and of their affinity to a certain drug, in order to provide relevant indications regarding a therapy ad hoc developed for each patient.

In this final section are presented the results of the early steps made in collaboration with the M. D. Anderson hospital of Houston. The collaboration with the hospital in this direction is still going on and a lot of work must still be done in this direction.

3.8.1. Early results

The group of Jonathan Trent at M. D. Anderson was interested in the creation of a benchmark of the mutations that mostly occur in the KIT chain during the treatment with the most used inhibitor drugs (the same mentioned in the previous section 3.7).

The role of molecular modelling in this framework can be extremely important. In fact, the creation of a real clinical benchmark regarding the effect of mutations on the affinity of KIT for a certain drugs would require an enormous amount of work and timings that are not compatible with the real necessity of the patients. In this direction modelling would allow to obtain a base benchmark in a slightly reduced timings, providing to the
medical doctors relevant information regarding “which inhibitor should guarantee better results in the situation of the patient”.

However, despite the powerful computational resources that are presently available, also from the point of view of the computer simulation the creation of such benchmark represents a non-trivial challenge.

Our clinical partners presented about one hundred of mutations occurring in the KIT. They were interested to simulate them with all the drugs previously mentioned. In this section we are going to present the early results achieved during the last year of phD activity in this direction. After an initial screening, the hospital highlighted the thirty most crucial mutations. Data shown in this sections refer to the simulations of the KIT receptor both in WT and mutated conditions in complex with imatinib.

Figure 3.22. The KIT-imatinib complex. In this representation, the KIT structure is represented as a gray-shadowed ribbon and the imatinib drug in red stick style. Water molecules, ions in solution and hydrogen atoms are not present for clarity.
Starting from the “wild type” (WT) complex, all the mutations were introduced with the same procedure used in the cases previously discussed. Each simulation run had the same length and used the same parameters and variables for the cases presented in the other sections. Energetic analyses were conducted with the MM-PBSA (see Material & Methods chapter for details) on a proper number of snapshot taken from the equilibrated phase of the dynamic trajectories of each complex of interest.

Table 3.9. Free energies of binding of KIT in complex with the imatinib inhibitors. Results are reported for the WT as well as for all the mutated cases. All energetic values are expressed in kcal mol\(^{-1}\).

<table>
<thead>
<tr>
<th>KIT complex</th>
<th>(\Delta G_{\text{bind}}) kcal mol(^{-1})</th>
<th>KIT complex</th>
<th>(\Delta G_{\text{bind}}) kcal mol(^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>-8.1</td>
<td>(\Delta 551-555)</td>
<td>-8.6</td>
</tr>
<tr>
<td>L576P</td>
<td>-7.8</td>
<td>(\Delta 552-553)</td>
<td>-7.9</td>
</tr>
<tr>
<td>W557G</td>
<td>-9.5</td>
<td>(\Delta 552-557)</td>
<td>-9.1</td>
</tr>
<tr>
<td>W557R</td>
<td>-8.8</td>
<td>(\Delta 553-558)</td>
<td>-9.1</td>
</tr>
<tr>
<td>V559A</td>
<td>-9.1</td>
<td>(\Delta 555-559)</td>
<td>-7.7</td>
</tr>
<tr>
<td>V559D</td>
<td>-9.4</td>
<td>(\Delta 556-557)</td>
<td>-7.8</td>
</tr>
<tr>
<td>V559G</td>
<td>-9.3</td>
<td>(\Delta 556-561)</td>
<td>-7.8</td>
</tr>
<tr>
<td>(\Delta 560)</td>
<td>-9.1</td>
<td>(\Delta 556-574)</td>
<td>-9.4</td>
</tr>
<tr>
<td>V560D</td>
<td>-8.1</td>
<td>(\Delta 557-558)</td>
<td>-8.8</td>
</tr>
<tr>
<td>V560E</td>
<td>-8.2</td>
<td>(\Delta 557-558+\Delta 560)</td>
<td>-8.8</td>
</tr>
<tr>
<td>V560G</td>
<td>-8.1</td>
<td>(\Delta 558-559+\Delta 565)</td>
<td>-9.0</td>
</tr>
<tr>
<td>(\Delta 576)</td>
<td>-7.8</td>
<td>(\Delta 559-560)</td>
<td>-9.1</td>
</tr>
<tr>
<td>(\Delta 579)</td>
<td>-8.1</td>
<td>(\Delta 559-565)</td>
<td>-8.3</td>
</tr>
<tr>
<td>(\Delta 550-558)</td>
<td>-8.8</td>
<td>(\Delta 564-576)</td>
<td>-7.7</td>
</tr>
<tr>
<td>(\Delta 550-559)</td>
<td>-8.6</td>
<td>(\Delta 570-576)</td>
<td>-8.8</td>
</tr>
</tbody>
</table>

It is worth noting that data collected in this table are consistent with the clinical evidences reported by M. D. Anderson. These were selected in a first time from a longer
list due to the possibility to have a clinical feedback to check the reliability of the molecular models used for this purpose. The models respect the trend reported by the hospital. This method can thus be considered as reliable as a starting point – many other mutations can be simulated and added to Table 3.9. The same mutated KIT chains may be simulated in complex with sorafenib, nilotinib, dasatinib and sunitinib, in order to complete the screenshot provided by this analysis.
Chapter 4

Conclusions

This chapter contains the final considerations on the results collected into this thesis. In this section we will summarize the main results achieved during the years of PhD activity and the future directions of the research will be indicated.
The notable progresses achieved during the past years in the genetic and medical biotechnology research allowed to identify the common genetic origin that is settled behind many different pathologies. Genes that, for unknown reasons, mutate during the lifetime of the patient can synthesize proteins and receptors that lose the control by grow factors. These aberrant proteins are consequently always active and phosphorylated, and give rise to a series of cascade signals that results in an uncontrolled growth of tumor cells.

In this framework the research based on “target therapy” represents a breakthrough – the design and the development of new inhibitor drugs that are able to penetrate into the deregulated tyrosine kinase domain, blocking the access to ATP, and deactivating the protein, became a key point in the treatment of these kind of tumoral diseases.

Even if these small molecules express an extraordinary efficiency toward tumors – side effects are also decreased if compared with the common therapies – the treatment of these pathologies is still far from the final solution. In fact, the patients treated with these inhibitors acquire almost unavoidably a sort of resistance toward the action of the drug, showing relapse and a new progression of the disease. What happens is that aberrant proteins (Bcr-Abl, PDGFRα) and receptors (KIT), synthesized by onco-genes, can present secondary mutations along their amino acidic chains. Even a single amino acidic substitution (or deletion and duplication) is able to affect so strongly the affinity between kinase and inhibitor molecule that the drug becomes ineffective.

The purpose of this thesis was to use molecular modelling to gain an insight into the binding between tyrosine kinases and determined inhibitor drugs.

In general, the works reported in this thesis highlighted that molecular simulation can be really considered as a powerful tool to support clinic scientists in the comprehension of drug-resistance phenomena. By creating a virtual (ideal) system, it is possible to analyze biological phenomena of extraordinary complexity from an privileged point of view.

In this thesis molecular simulation was used to gain unique details into the factors that drive and favor the complexation between drugs and proteins. We demonstrated that our molecular models are able to reproduce with optimal reliability the experimental
evidences of drug resistance. Our data showed that the appearance of even a single mutation along the amino acidic chain of protein and receptor can affect drastically the thermodynamic energetic stability of binding generating these resistance effects.

In this framework, we were able to highlight how different drugs, with different functionality (expressed in terms of different penetration of the drug within the binding site of kinases) can be effective or ineffective in the binding depending on the situation of the patient. This pointed out another important achievement – with modelling techniques it was possible to trace the direct relationship that exists between structure, functionality and efficiency of a certain drug.

In addition, we were able to divide the most widely used drugs into two different groups, based on the different position occupied in the binding pocket of proteins and receptors during the inhibition – as said, the different functionality of these drugs results in a different efficiency during the inhibition. For example, in the case of the KIT-L576P mutation, that takes place in the juxtamembrane zone of the KIT receptor, drugs that penetrate deeper in the binding pocket lose their efficiency. Our results identified the cause in the rearrangement that interest the deep zone of the binding site due to the presence of this mutation. The consequent adjustment creates conflicts (in terms of thermodynamic stability) with drugs that penetrate deeper. On the contrary, inhibitors that stop in the external zone of the protein (dasatinib and sunitinib) are able to preserve their efficiency.

The fact that each drug can be efficient or ineffective depending on the mutations expressed by each patient thus appears to be an immediate consequence.

This important result suggests that the deep comprehension of the binding mechanism provided by in silico approaches is the first step toward the conception of novel drugs or toward the optimization of the ones that are already available in the market.

More important, data suggest that the future efforts should be addressed to the creation of a benchmark reporting the responses of the most used drugs toward the mutations that most often occur in the protein and receptor structures characteristic of these kind of diseases. In this thesis we reported the early results related to the imatinib drug complexes, but a lot of work must still be done in this direction.
In conclusion, molecular modelling can really support the cancer research by acting as a virtual microscope, able to provide unique insight into the binding mechanism between proteins and inhibitor drugs and into drug resistance phenomena.
References


45. Case DA, Darden TA, Cheatham TE, et al. AMBER 9, University of California, San Francisco, CA USA, **2006**.


References


Ringraziamenti

Voglio ringraziare tutte le persone che mi hanno supportato durante l’intero periodo del dottorato. Prima di tutto, il “grazie” più grande va a chi mi ha incoraggiato ad intraprendere questo percorso. Ringrazio la mia famiglia che mi è stata vicino in tutti questi anni. Questa tesi è dedicata a mia madre e a mio padre, che mi hanno sempre insegnato l’importanza della dedizione, dell’impegno e della forza di volontà per migliorararsi, e cercare di raggiungere i propri sogni. Voglio ringraziare e abbracciare mio fratello Matteo, mia sorella Mary, la mia Tere e tutti gli amici che sono sempre stati con me e mi hanno sempre appoggiato. Ringrazio il mio supervisore Sabrina, per avermi accompagnato durante questi anni di dottorato e per avermi insegnato molte cose che saranno fondamentali per il mio lavoro futuro, e anche Maurizio, per il suo esempio di serietà e professionalità. Devo molto ad entrambi per avermi dato l’opportunità di muovere i primi passi nel mondo della ricerca.