Mapping The Kinase Mutation Landscapes: A Computational Exploration

Vilacha, Juliana Fátima

DOI:
10.33612/diss.909182162

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2024

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
INTRODUCTION

Hoping for the best, 
prepared for the worst, 
and unsurprised by anything in between

Maya Angelou

Parts of this chapter have been published in:
ABSTRACT

The parallel advances of different scientific fields provide a contemporary scenario where collaboration is not a differential, but a requirement. In this context, crystallography has had a major contribution to the medical sciences, providing a “face” for targets of diseases that previously were known solely by name or sequence. Worldwide, cancer still leads the number of annual deaths, with 9.6 million associated deaths, with a major contribution from lung cancer and its 1.7 million deaths. Since the relationship between cancer and kinases was unraveled, these proteins have been extensively explored and became associated with drugs that later attained blockbuster status. Crystallographic structures of kinases related to lung cancer and their developed and marketed drugs provided insight into their conformation in the absence or presence of small molecules. Notwithstanding, these structures were also of service once the initially highly successful drugs started to lose their effectiveness in the emergence of mutations. This chapter focuses on a subclassification of lung cancer, non-small cell lung cancer (NSCLC), and major oncogenic driver mutations in kinases. We review how crystallographic structures can be used, not only to provide awareness of the function and inhibition of these mutations but also how these structures can be used in further computational studies aiming at addressing these novel mutations.

Keywords: cancer; NSCLC; mutation; kinase; EGFR; ROS1; X-ray crystallography; molecular modeling
1.1. Introduction

Cancer is a general term to define a myriad of medical conditions that can affect different tissues in the body. A common characteristic is the abnormal growth of a cell that later develops the ability to spread to other tissues, in a process known as metastasis [3]. Worldwide, it is the second major cause of death, responsible for one in six deaths. Within cancer-related mortality, lung cancer is responsible for more than one million deaths annually - populating the top of the list of deadliest cancer types, a situation that is likely to increase. Lung cancer can be subdivided into Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC), with the latter being diagnosed in around 85% of lung cancer patients [4]. Advances in molecular diagnostic techniques unveiled details of genes acting as drivers in different disorders identifying a family/class of proteins responsible for controlling key features of cell development: the kinase family [5].

Genetic events, such as gene amplification, activating mutations, and fusions or chromosomal rearrangement were found in druggable kinases, including the Epidermal Growth Factor Receptor (EGFR) (30%), Anaplastic Lymphoma Kinase (ALK) (10%), rapidly accelerated fibrosarcoma isoform B (BRAF) (1.7%) among others (>1%) [6, 7]. The Kirsten rat sarcoma viral oncogene homolog (KRAS) is also a major oncotarget for lung cancer with the highest incidence (40%) in either smokers or non-smokers. This GTPase was often overlooked due to its “undruggability”. However, new results show the potential for a new class of KRAS targeting small molecules [8].

The first kinase structure was published in 1991, in which researchers from the University of California, San Diego solved the structure of the cyclic-AMP dependent protein kinase (PKA), a kinase with the ability to phosphorylate tyrosine side chain in substrates. This structure opened the doors for what would be one of the most explored fields during the new millennium; the pharmacological assessment of kinase proteins [9]. During the early 2000s, the booming field brought numerous kinase inhibitors to the clinic, with imatinib, an inhibitor designed for Philadelphia chromosome-positive chronic myelogenous leukemias, at the vanguard [10]. With the NSCLC-related kinome unraveled, there was an emerging need for the characterization of these targets, not solely to understand their molecular mechanisms but also for rational drug development.

In parallel with the advances in the medical sciences and structural biology techniques, another related field made huge strides over the years, namely molecular modeling (MM). MM comprises theoretic and computational methods for representing, mimicking, and manipulating molecules: from small (water) to larger structures, such as cellular proteins and membranes or even simple cells [11, 12]. Molecular modeling can analyze atoms or descend further into quantum chemistry [13]. In drug design, the mechanist approach is the most common due to its ability to describe a diversity of molecules, from water to oligomeric protein complexes, at the level of atoms, bonds, and angles, at high accuracy, in a relatively short time [14]. This chapter will focus on homology modeling, molecular docking, molecular dynamics, and free energy calculations involving selected kinases implicated in NSCLC.

Homology modeling (HM) is a bioinformatic tool commonly used to obtain the three-dimensional structures of proteins that are so far experimentally unresolved. Unlike the
1. Introduction

Experimental elucidation of protein structures that might be delayed for difficulties with protein expression and further crystallization, HM is based on using proteins with a high level of sequence homology as a template for the desired target. Like any technique, homology modeling has limitations, in this case, it is necessary to have a high-quality template, optimally another protein from the same family, with a high level of sequence similarity (> 40%) [15]. HM is also limited in the prediction of highly flexible motifs such as loops or tails. Common consensus classifies the homology model as a low-resolution structure [16]. In addition to contributing to providing structural models of novel structures, HM has also been extensively used to generate three-dimensional models of mutants in the absence of experimental structural information [17].

With a target structure in hand, it is possible to search for potential binding pockets, clefts, or cavities in the structure that would provide anchoring points for other molecules. Tools such as FTmap, AnchorQuery, and PocketQuery are a few of the hundreds that are used to map and recognize druggable clefts on proteins [18–20]. A more traditional manner of targeting macromolecules is to study the site in which native ligands bind and aim to design a competitive molecule. Molecular docking is a useful tool for probing libraries of molecules or chemical fragments in a desired binding pocket [21]. The prototype molecule conformation is accessed through the exploration of a large conformational space representing various potential binding modes and rank-ordering by affinity. These steps are performed in a cyclic process by analyzing each ligand conformation with the selected scoring function, thereby converging to a minimum energy solution. During the conformational search step, the ligand undergoes modification on its torsional (dihedral), translational, and rotational degrees of freedom. This method explores the energy landscape of the conformational space and converges them into the most likely binding mode with the minimum energy [22]. However, molecular docking is not limited to protein-small molecules studies, it is also a useful tool to explore protein-protein and DNA-ligand interactions [23, 24].

Although homology modeling and molecular docking are useful tools, they lack an appropriate assessment of protein dynamics. To address this matter, Molecular Dynamics (MD) simulations are an option. MD is based on the computational simulation of molecules, considering the physical motion of atoms. In this methodology, each atom of the whole protein structure has its position and velocity determined through Newton’s equations of motion. The first step is to portray the system of interest with the selected force field while considering the temperature and pressure for the simulation. Secondly, by computation of the forces acting on each atom of the whole complex, it is possible to obtain the position and velocity of these atoms at a certain moment. This cycle is repeated through a determined period, specific for each experiment. The final result is the trajectory and progression in time of the simulated system [25].

As the mass of every atom in the system is known, only the force is required to be calculated to obtain the acceleration. In MD, this force is determined by Force Fields such as AMBER, CHARM, NOVA, and YASARA. These force fields determine the force acting on the system using molecular interaction potentials, which can be established with the use of quantum chemistry calculations or experimental data aiming to compute how each type of interaction contributed to the global function, and thus to the whole system [26]. Force field equations consider intramolecular forces, such as the bond between
atoms (often considered as “springs”), angles between bonds and dihedrals, as well as intermolecular forces, such as Van der Waals and electrostatic interactions [27]. In this methodology, solvent layers are also computed and water molecules can be described through different models [28–30]. MD simulations have massively improved and are an invaluable tool in modern science [31].

The remainder of this chapter is organized as follows; first, we provide an overview of the major structural features of the kinase family. Then I provide an in-depth description of two major kinases involved in NSCLC, EGFR and ROS1, with a focus on available x-ray crystallography structure and their interaction with small molecules. I will also provide an overview of the use of crystal structures in molecular modeling studies. Finally, I provide the aims and outline of the rest of the thesis.

1.2. Kinases: A structural overview

The involvement of kinase proteins in key regulatory aspects of cell biology is powered by their ability to modulate other proteins through a phosphorylation reaction, where the γ-phosphate group of Adenosine Triphosphate (ATP) molecules is transferred to selected amino acids of a substrate. Phosphorylation is the most common protein modification in signal transmission, mostly due to its reversible nature by dephosphorylation performed by phosphatases [32]. The nature of the phosphorylated residue guides the classification of kinase proteins into serine/threonine kinases or tyrosine kinases. However, a small group of proteins can target both threonine and tyrosine amino acids (dual specificity kinases) and are exemplified by the Dual Specificity Mitogen-Activated Protein Kinase 1 (MEK1) and 2 (MEK2) [32, 33].

The structure of the kinase domain can be generically depicted as a bilobal structure with a larger C-terminal lobe, presenting several conserved α-helices and short β-strands, connected through a hinge to a smaller N-terminal lobe, a composition of a five-stranded anti-parallel β-sheet (β1–β5) and a roving α-helix (αC-helix). In Figure 1.1, the crystal structure of the Hepatocyte Growth Factor Receptor (HGFR or c-MET) in the presence of ATP is used to exemplify the general kinase domain folding (Protein Data Bank (PDB): ID 3DKC). The N-terminal presents a conserved glycine-rich (GxGxΦG) loop, occurring between the β1- and β2-strands, responsible for positioning the β- and γ-phosphate groups from the ATP molecule for catalysis. The glycine-rich loop is also known as the G-loop or P-loop, with the latter being the most common in kinase-related literature. The β1- and β2 strands also harbor the adenine moiety of ATP, contributing to its stabilization [34].

Within the N-terminal lobe, a characteristic interaction is often observed involving a conserved lysine from the β3-strand and a glutamate residue occurring in the αC-helix, this salt bridge is a precondition for the active state. The presence or absence of such interaction is based on the positioning of the roving αC-helix. In the kinases, in this thesis, once the αC-helix is orientated toward the ATP binding pocket and the salt bridge is present it can be classified as αC-in. The outward positioning of the helix and absence of the lysine-glutamate interaction is known as the αC-out [35]. The rotation of the αC-helix is often used in MD studies to classify the resultant structures into either active or inactive [36].
1. Introduction

Figure 1.1 | The sequences of the kinases focused on in this thesis are represented with relevant residues highlighted. (A) Epidermal Growth Factor Receptor (EGFR) (UniProt ID: P00533), (B) Proto-Oncogene c-ROS1 (ROS1) (UniProt ID: P08922). (C) c-MET in complex with ATP (PDB: ID 3DKC) is used as a general representation of a kinase domain with the C-terminal lobe colored blue with the hinge motif colored pink, N-terminal lobe colored yellow with the P-loop residues backbone depicted as sticks and the regulatory αC-helix colored in orange. Nitrogen and oxygen atoms are colored blue and red, respectively. The ATP molecule is depicted with carbon atoms in green and phosphate atoms in orange, magnesium ion is depicted as a green sphere.

Although the αC-helix positioning and salt-bridge presence are necessary for the full activation of the protein, they are not sufficient. The C-terminal lobe contains a flexible segment, the activation segment, or A-loop, a motif initiated by a conserved sequence of Asp-Phe-Gly (DFG) whose rotation has major effects on ATP binding pocket occupancy. Once the kinase is in an active conformation, the side chain of the aspartate residue occupies the ATP binding pocket (DFG-in), thereby coordinating a magnesium ion [37, 38].

Once the crystal structure of inactive c-Src tyrosine kinase was elucidated, a different DFG conformation was observed [39]. The aspartate side chain was flipped 180° when compared to the active conformation, and swapping positions with the direct neighbor residue, phenylalanine, characterizing a DFG-out conformation. Once the DFG-out state is formed, an allosteric pocket is formed contiguous to the ATP binding area. Guarding the new pocket there is a “gatekeeper” residue, a hotspot for resistance mutation against inhibitors in multiple kinases that will be addressed later in this chapter [30, 38].

The EGFR kinase domain follows the aforementioned description with specific se-
quences for the mentioned motif represented in the highlights in Figure 1.1. However, despite the salt bridge and DFG-in conformation being present in the few available crystal structures, there is an urge for further analysis of this kinase’s dynamics/behavior. The highlighted structural motifs featured in Figure 1.1 are dynamic and have been extensively used not only for kinase state identification but also for the design of inhibitors. Due to their contribution to the proteins’ plasticity, they are commonly analyzed for insights into the behavior of a protein of interest.

1.3. THE EPIDERMAL GROWTH FACTOR RECEPTOR (EGFR)

The Epidermal Growth Factor Receptor (EGFR), also known as erbB1 or HER1 (UniProt ID: P00533), is a member of the erbB family of receptor tyrosine kinases (RTKs). Structurally, members of the erbB family (erbB 1-4) consist of an extracellular domain (ectodomain), a membrane-spanning domain, and an intracellular domain. The extracellular domain of the erbB family members is a target for a variety of molecules including the Epidermal Growth Factor (EGF), epiregulin (EPR), neuregulin (NRG family), and the transforming growth factor-α (TGF-α) [40, 41]. In the absence of extracellular stimulus, all four receptors are found inactive in the cellular membrane, forming homo- or heterodimers upon ligand binding [42]. The intracellular region is further divided into a tyrosine kinase domain (residues 696–976), and a C-terminal regulatory region (residues 977–1210). The tyrosine kinase domain is responsible for the catalytic activity—conversion of an ATP molecule into ADP through cleavage of the bond between the γ and β phosphate groups and the release of a phosphate [40].

The erbB family, in a typical cellular setting, is responsible for translating the external stimulus from ligands into intracellular signaling as depicted in Figure 1.2. Upon activation of a tyrosine kinase receptor, such as erbB family members, the complex Grb2/SOS is recruited for further binding to inactive Ras proteins, promoting the release of GDP and binding of GTP [43]. The activation of downstream pathways is associated with cell growth and cell survival [44]. Consequently, the aberrant activation of these family members is often linked with a variety of human cancers, most notably NSCLC [45]. EGFR mutations are one of the major causes of NSCLC formation and progression and appear more frequently in never or light smokers, women, and East Asian NSCLC patients [46]. In a healthy cell, the absence of extracellular stimuli drives EGFR monomers into an auto-inhibitory, tethered conformation in which the dimerization arm is buried [47, 48]. Activating ligands bind bivalently on EGFR and trigger a large conformational change in the extracellular domain, in which the dimerization arm, a β hairpin-like motif, becomes exposed to the aqueous environment and is thus able to dimerize [49]. Recent reports show the presence of a mixed population of inactive monomers and homodimers in the cellular membrane. In the inactive dimer, the monomers adopt a symmetric configuration and are kept together through interactions of the intracellular domain, the transmembrane domain, and the C-terminal tail of the extracellular domain. In response to ligand binding, a conformational change leads to a rotation of the transmembrane and intracellular domains of the receptor, yielding an active, asymmetric dimer [50].

Elucidation of EGFR kinase domain structure complies with the canonical structure previously mentioned, as seen in Figure 1.1, with representations of the α-helix rich C-
1. INTRODUCTION

Figure 1.2 | Summary of the main signaling pathways involving the erbB family, Echinoderm microtubule-associated protein-like 4 (EML4)-ALK fusion protein and Rapidly Accelerated Fibrosarcoma (RAF) family. Rat sarcoma (RAS) protein family implication is also represented in either GDP- or GTP-bound states. These signaling pathways regulate and control important downstream cellular functions such as apoptosis, cell growth, survival, angiogenesis, and migration. Additional kinases represented in this figure are the Dual Specificity Mitogen-Activated Protein Kinase (MEK), extracellular signal-regulated kinases (ERK), and the protein kinase B (also known as AKT).

terminal lobe and an N-terminal lobe containing a five stranded \( \beta \)-sheet, its P-loop, and the mobile \( \alpha \)C-helix [51]. As shown in Figure 1.3, the plasticity of EGFR can be measured between its active and inactive conformation by comparison of the specific motifs such as the clear coiled conformation of the activation segment in the inactive form against an elongated loop in the active EGFR, as an example [27, 52, 53].

In analyzing the binding of ATP and its analogs to EGFR, the aromatic N1 nitrogen atom serves as a hydrogen bond acceptor for the backbone amino group from M793 while the N6 amino group serves as a hydrogen bond donor for Q791 (PDB: ID 2GS6) [52, 54]. A combination of MD simulations with a Molecular Mechanics Generalized Born Surface Area (MMGBSA) method was employed to elucidate the specific structural elements that stabilize ATP binding in both active and inactive conformations. MMGBSA allows for the assessment of binding free energy of protein-ligand complexes with a modest computational input but with proven contributions towards a higher quality evaluation of small ligands binding to biomacromolecules [55]. In the EGFR kinase domain, ATP binding is stabilized by the formation of hydrogen bonds and salt bridges between the
negatively charged phosphate group and residues K745 and R841, in both the active and inactive states throughout the simulations. Furthermore, an ionic bond is formed between the negatively charged phosphate and Mg\(^{2+}\), which is correctly orientated by D855 and N842 [56, 57].

In addition to the conformational changes required for EGFR activation, a dimerization process is also necessary. As previously mentioned, inactive EGFR is often found as symmetric dimers while activation is characterized by an asymmetric complex of two monomeric units. In the case of an EGFR-EGFR dimer, one of the units acts as an active partner contributing with its N-terminal while the other is the passive partner, contributing with its C-terminal. The dimer interface is governed by hydrophobic interactions involving residues L704, I706, L760, L782, and V786 from the active partner, and I941, Y944, M945, V948, and M952 from the receiver kinase unit contributing with its C-lobe. The asymmetric dimer formation results in the transphosphorylation between the kinase domains and their activation [52].

Early attempts in crystallization of the monomeric inactive EGFR TK domain were challenging due to the spontaneous dimerization of EGFR at high concentrations. Wood et al. elucidated the first inactive structure in the presence of EGFR inhibitor lapatinib.
Lapatinib is an ATP-competitive drug with a 4-anilinoquinazoline scaffold designed to interact with the hinge motif (PDB: ID 1XKK). Lapatinib explores the allosteric pocket guarded by the gatekeeper residue with an extended moiety using the hydrophobicity of phenolic rings substituted with chlorine and fluorine atoms to form Van der Waals interactions as depicted in Figure 1.4 [59]. The inactive conformation closely resembles the structures of inactive Src and Cyclin-Dependent Kinase (CDK) proteins, mainly due to the positioning of the αC-helix [58].

![Figure 1.4](image.png)

Figure 1.4 | Two-dimensional representation of lapatinib and first-generation inhibitors, gefitinib and erlotinib. The 4-aminoquinoline scaffold, common to all drugs, engages in hydrogen bonds with hinge residue M793 while hydrophobic substituents contribute to van de Waals interactions with residues from the allosteric pocket, such as L788, V766, and L858. Interactions were analyzed and plotted using Discovery Studio Visualizer. Atoms color scheme: carbon(grey), nitrogen (blue), oxygen (red), sulfur (beige), chlorine (green), and fluorine (light blue).

A more elegant methodology with the incorporation of the V948R point mutation has been employed to study the inactive TK domain in which the ATP nucleotide analog AMP-PNP is coupled (PDB: ID 2GS7). This mutation takes place in the dimer interface and disrupts the hydrophobic interactions between the active and the receiver units by introducing the polar residue arginine. Consequently, the mutation restricts the dimerization of erbB members and allows for studies of monomeric wild-type (WT)
1.3. THE EPIDERMAL GROWTH FACTOR RECEPTOR (EGFR)

The first crystallographic structure obtained in the active state was WT EGFR in the apo form (PDB: ID 1M14). In the active EGFR TK domain conformation, the αC-helix becomes highly structured, adopting an αC-in conformation allowing the formation of the catalytically important salt bridge (E762-K745). Furthermore, the A-loop is positioned away from the active site, allowing substrate access to the catalytic site due to the DFG-in state.

The increased availability of EGFR structures contributed to more in-depth studies to elucidate the dynamics of EGFR, both as a monomer and as a member of the activated dimer. A set of 25 EGFR crystallographic structures was used to build an ensemble structure that would be further submitted to molecular dynamic simulations and analyzed as standalone kinase domains or as active/receiver components of an EGFR homodimer. This extensive study showed that asymmetric dimerization might not only stabilize the active conformation, thus organizing the ATP binding site, but also affect regions distal from the dimerization interface such as the A-loop. It is relevant to highlight concerns about the duration of the simulations, here consisting of 100 ns, which are potentially not enough to accurately model the intrinsically disordered intermediate between the active and inactive conformations. Songtawee et al. show that MD simulations can sample different experimentally obtained conformations by comparing the output of MD simulations with the aforementioned structures, the high level of similarity between MD outputs and crystallization data strengthens the reliability of computational methods such as molecular dynamics to explore kinase dynamics.

In the catalog of somatic mutations in cancer (COSMIC) database, more than 594 types of mutations have been registered in the EGFR TK domain, of which the majority (93%) are located in exons 18, 19, 20 and 21. The most common class I activating mutations are deletions in exon 19, spanning through residues 746–750, accounting for approximately 44% of all EGFR mutations. Class II mutations comprise activating, and resistance-acquiring point mutations. Classical variants are exemplified by the activating L858R (39.8%), the resistance acquiring T790M, and the rare sensitizing mutation G719A/C/D/S (3%). Class III mutations include exon 20 in-frame insertions and duplications.

The two most commonly found activating mutations, L858R and exon19 deletions account for over 90% of all sensitizing mutations and thus have been termed “classical” activating mutations. L858 is located towards the N-terminus of the activation segment, directly neighboring the DFG motif. When in the inactive conformation the L858 is part of a hydrophobic cluster composed of hydrophobic and/or aromatic residues, F723, L747, M766, and L788, responsible for stabilizing the coiled stated of the activation segment. Structures of L858R mutants are available with a variety of ligands, from ATP analogs (PDB: ID 2EB3) to natural products (PDB: ID 2ITU). The multiple ligand complexes all share the same active conformation with an uncoiled A-loop and αC helix-in conformation.

The L858R point mutation is characterized by a 10 to 100-fold increase in the affinity for TKIs. A structural comparison of WT/AMPPNP (PDB: ID 3VJO) with L858R/AMPPNP (PDB: ID 2EB3) identified that the side chain of F723 protrudes outwards towards the active site and interacts with R748. This enlarges the active site cleft, which allows faster release of ATP, while also making the active site more attainable for TKIs.
Further information on the L858R mutation was provided by Ding et al. through long-duration (500ns or more) atomistic MD simulations, and experimental data. The MD studies provided evidence on the process of switching from an inactive state into an active conformation. Simulations involving gefitinib bound to either the WT or L858R show higher binding energy to the active than to the inactive conformation. This data explains why inhibitors such as gefitinib are more successful in the presence of activating mutations than in gene amplifications in the treatment of EGFR-positive NSCLC patients. Since L858R drives the kinase domain into an active conformation, disturbing the equilibrium between active and inactive, it contributes to drug binding by providing a more accessible (open) conformation for the drug [70].

Deletions on exon19, specifically the common Δ746ELREA750 (Del19), occur in the N-terminal loop, between the β3-strand and the αC-helix. MD studies performed by Tamirat et al. on the inactive and active forms of the Del19 mutant found that the active state is favored due to the stabilization of the E762-K745 salt bridge. This results from a decrease in the β3-αC loop flexibility, which stabilizes the αC-helix in the active αC-in conformation. Furthermore, the Del19 causes an inwards conformational shift of the αC-helix, which disrupts the hydrophobic cluster in between the αC-helix, thereby promoting the activation of the TK domain [71]. Similarly to L858R point-mutants, Del19 EGFR mutants display a decreased affinity for ATP binding (KM) and a lower Ki for first-generation TKIs such as erlotinib [71, 72]. However, exon 19 deletions and consequent residue insertions display large heterogeneity and thus exhibit differential drug sensitivity [73, 74].

Upon identification of the aforementioned mutations as major biomarkers of NSCL cancer, there has been a keen interest in the development of drugs able to inhibit the enhanced kinase activity provided by these mutations. Efforts culminated with the development of first-generation inhibitors gefitinib and erlotinib, both reversible ATP competitive inhibitors sharing a 4-anilinoquinalzoline scaffold [75]. As pictured in Figure 1.4, both drugs are capable of interacting with M793 similarly to the adenosine ring of ATP, through a hydrogen bond with the residue backbone. The presence of the 3-chloro-4-fluoro aniline allows gefitinib to explore the allosteric hydrophobic pocket through interactions with L788, and T790. The methoxy moiety is within Van der Waals contact of G796 (PDB: ID 4WKQ). The 6-propyl morpholine ring on gefitinib extends into an area exposed to the solvent and was implemented to improve pharmacokinetic properties [76]. Compellingly, gefitinib has also been identified to bind in a second conformation to the L858R mutant. The second conformation exhibits a 180° rotation of the aniline ring, which allows the chloride substituent to interact with the sidechain of R855 via a halogen bond through the coordination of a water molecule. In both cases, the ether group extends outwards from the ATP binding pocket towards the aqueous environment [51].

In the active EGFR conformation, the erlotinib's anilinoquinazoline ring is stabilized by seven hydrophobic (L718, A743, L788, L792, P794, and L844), three polar (T790, Q791, and T854) and three charged residues (E762, K45, and D855), while the solvent-exposed substituents interact with F795 and G796 (PDB: ID 1M17) [77]. In the inactive state, erlotinib is stabilized by the seven hydrophobic interactions stabilized in the active state and by an extra hydrophobic interaction from V726. Furthermore, it is stabilized by the same three polar residues and by three charged residues (K745, D800, and D855) [78].
1.3. The Epidermal Growth Factor Receptor (EGFR)

Initial studies suggested that both drugs recognize the active conformation of EGFR. However, erlotinib can bind to both active and inactive, conformations [79]. Crystallographic studies were able to co-crystallize the inactive TK domain in complex with erlotinib (PDB: ID 4HJO) providing further crystallographic evidence that erlotinib can bind to both states [79].

Erlotinib and gefitinib both showed greater potency for L858R EGFR against WT EGFR. Notably, gefitinib binding to EGFR L858R mutant is increased 20 folds when compared to WT EGFR and thus, preferentially inhibits L858R positive cancer cells, leading to their consequent apoptosis and cancer remission while sparing healthy cells [51]. Afatinib, a second-generation approved in 2013, is an irreversible ATP competitive anilinoquinazoline, which harbors an acrylamide reactive group as shown in Figure 1.5. Crystallographic studies by Solca et al. on the WT EGFR in complex with afatinib, (PDB 4G5J) showed a hydrogen bond between M793 and its core quinazoline ring but, most importantly, the electron density map identified a covalent bond formed between C797 and the acrylamide group in the active state of the kinase [80].

Figure 1.5 | Two-dimensional representation of second-generation inhibitors, afatinib and dacomitinib, and third-generation inhibitor, osimertinib. Although all three drugs follow a similar binding mode to first-generation inhibitors through hinge-binding scaffolds, osimertinib lacks interactions with the allosteric pocket. The proximity of the drugs’ warhead to C797, its covalent bond partner, is also depicted. The color scheme is described in Figure 1.4.

G719X (where X can be alanine, aspartic acid, cysteine, or serine) is a rare sensitizing
point mutation on exon 18 accounting for approximately 3% of all EGFR TK domain mutations, with G719S (PDB 2EB2) being the most common variant [67, 81]. G719 is located on the P-loop connecting strands 1 and 2, contributing to the stabilization of ATP phosphate groups. Furthermore, G719 is part of the hydrophobic cluster found during the inactive conformation and generates a steric hindrance that helps position the αC-helix from the active site in an αC-out conformation. The structure of the G719S mutant in complex with AMP-PNP (PDB: ID 2ITN), gefitinib (PDB: ID 2ITO), and a staurosporine analog (PDB: ID 2ITQ) are available. Unlike L858R and del19 mutations, the G719X mutation does not promote receptor dimerization but instead influences intrinsic structural components favoring receptor activation [72].

The presence of a serine residue at position 719, shows to be sensitive to gefitinib with an inhibitory concentration (IC\textsubscript{50}) of 0.18 \(\mu\)M against 1.04 \(\mu\)M found for the WT EGFR. However, the presence of a secondary mutation at position 790 (T790M) increases the IC\textsubscript{50} by 10-fold (IC\textsubscript{50} = 1.86 \(\mu\)M). Interestingly, analysis of the constant of dissociation (K\textsubscript{d}) for the double mutant (K\textsubscript{d} = 5.6 nM) shows a tighter binding of gefitinib when compared to either the single mutant (K\textsubscript{d} = 31.9 nM) or the WT (K\textsubscript{d} 14.2 nM), indicating that double mutant diminished affinity for gefitinib is not due to reduced drug binding. However, a plausible explanation is raised by the kinetic studies, which show a ratio between the kinase activity (k\textsubscript{cat}) and the Michaelis–Menten constant (K\textsubscript{m}) comparable to the WT, indicating that the mutation T790M restores the nucleotide-binding ability [67].

Treatment of L858R, Del19, and G719X with first- and second-generation TKIs shows improved overall survival when compared to the classical chemotherapy [82, 83]. However, after a median of nine to thirteen months, EGFR-positive NSCLC treated with first- or second-generation typically acquire the resistance mutation T790M [84, 85]. The T790M resistance mutation is analogous to the imatinib-resistant bar-ABL fusion harboring the T315I mutation and it accounts for more than 50% of all EGFR TKI-resistant mutations [86].

T790M is referred to as the gatekeeper mutation and it is located at the back of the ATP-binding site [87]. Just like L858R, the T790M mutation stabilizes the active conformation of the TK domain, but via a different mechanism as shown by free energy calculations. M790 is part of a hydrophobic cluster formed in the back of the ATP binding site of the N-lobe and interacts with M766, located on the αC-helix. This hydrophobic interaction further extends towards the F856 and the catalytically relevant DFG motif [52, 88]. Once the methionine replaces the threonine residue, the αC-in conformation is stabilized.

Thermodynamic integration (TI) analysis is a theoretical method able to correlate free energy divergence between two given states of a system even in different spatial coordinates arising from long MD simulations [89]. Park and colleagues’ combination of TI with MD simulations showed that the T790M stabilizes the active (αC-in) and intermediate disordered form of active apo EGFR while disfavoring the inactive. However, it does not repress αC-intrinsic disorder and, consequently, is not believed to favor dimerization. Combination of MD simulation with MMGBSA starting from both active and inactive conformations, in the presence of the T790M alone or combined of L858R shows that only erlotinib binding energy is decreased by the gatekeeper mutation while lapatinib is not affected [90].

The rise of resistance to first and second-generation TKIs by acquiring T790M is
believed to emerge due to changes in the stability of ATP and drug binding [91]. It has recently been proposed that T790M resistance is a consequence of the restoration of ATP sensitivity similar to that of WT EGFR [56, 92]. Comparisons of the L858R mutant with the WT EGFR showed that the variant amplifies the conformational landscape of the kinase domain. Interestingly, the co-existence of L858R with T790M as a double mutant presents a conformational landscape like the WT EGFR. Such a similar conformational profile is associated with the restored ATP affinity of the double mutant being comparable to the WT [56].

Resistance also stems from steric hindrance clashes from the replacement of threonine by methionine within the ATP binding pocket, which contributes to the reduced binding of reversible TKIs [87]. However, more recent studies show that both gefitinib and erlotinib retain low nanomolar binding affinity towards the T790M mutant, proving that drug binding is still possible despite being limited [88]. Limited binding affinity might still correlate with response to a drug yet, clinical assessment of these results might lead to a decision to withdraw the drug since the advantage of targeting the mutated kinase rather than its wild type is lost. The therapeutic window for oncology drugs is a major point in medical decision-making [93].

Following the observation that first-generation drugs retain binding affinity for the T790M mutant free-energy calculations demonstrated that gefitinib binding to the T790M and L858R are more energetically favorable than binding to WT EGFR. In addition, MD studies have shown that gefitinib binding alters the conformation of the αC-helix whilst the activation loop maintains an active conformation. Overall, it has been observed that the T790M mutation does not ablate gefitinib binding as experimentally demonstrated by Gajiwala et al. through in vitro phosphorylation analysis of T790M in the presence of L858R. Besides, crystal structures of the complex of gefitinib in the presence of the gatekeeper mutation (PDB: IDs 3UG2, 4I22) [91, 92].

The emergence of T790M can also follow the primary G719X activating mutation, leading to a double mutant with synergistic interactions that stabilize the active conformation. Analysis of the gefitinib-double mutant complex (PDB: ID 3UG2) showed that gefitinib binds to the double mutant like the WT EGFR [67]. Intriguingly, gefitinib binds the double mutant 6-fold stronger ($K_d = 5.6\text{nM}$) than the single G719S mutant ($K_d = 31.9\text{nM}$) providing evidence that binding is still possible and is not hampered via steric hindrance. However, the G719S does become 10-fold less sensitive to gefitinib when acquiring the T790M resistance mutation. When in the presence of AMPPNP (PDB: ID 3VJN), the methionine in position 790 forms a more stable structure with AMPPNP when compared to the WT. An important observation for future drug development is that the double mutant also decreases the size of the hydrophobic cleft formed between L718 and G796, and therefore future drug prototypes that aim to treat the T790M mutant should avoid this cleft [67].

Another piece of evidence proving that steric hindrance does not ablate drug binding is that afatinib has been co-crystallized in the active receptor conformation in the presence of the T790M mutation (PDB: ID 4G5P) [80]. In vitro kinase assays identified that afatinib has a 100-fold higher potency against the L858R/T790M double mutant when compared to gefitinib [80]. However, the concentrations necessary to bring an inhibitory effect to the T790M point mutation might not be attained in the clinic under standard dosing
regimens [94].

Due to the need to find a viable treatment option for the emergence of the resistant T790M mutation, third-generation TKIs have been developed. Osimertinib, whose molecular structure is disclosed in Figure 1.5, is an irreversible EGFR inhibitor comprising a 2,4-diarylaminopyrimidine scaffold utilized for the L858R/T790M or exon19deletion/T790M mutants and shows a 200-fold preference for the double mutants over the WT EGFR. A combination of the information provided by the crystal structure of osimertinib with WT EGFR (PDB: ID 4ZAU) and homology modeling helped to elucidate the binding mode of osimertinib to the mutated TK domain [95].

Yosatmadia et al. modeled osimertinib binding using the previously known crystal structure of T790M EGFR in complex with dacomitinib (PDB: ID 4I24) [91, 95]. The L858R, T790M, and L858R/T790M mutations do not directly contribute to osimertinib binding but do favorably alter the TK domain conformation and dynamics that enhance the drug binding [95]. Osimertinib engages in a hydrogen bond with the M793 backbone while Van der Waals interactions contribute to the drug orientation within the pocket. Specifically, the phenyl ring sits in a hydrophobic sandwich between L718 and G796, and the methyl-indole moiety is within Van der Waals distance from G719, F723, and V725. The interactions from the indole ring, especially with the aromatic side chain of F723, are believed to be responsible for the positioning of the P-loop towards the ATP binding pocket. As an irreversible ligand, osimertinib is capable of engaging in a covalent bond with C797 [95].

Unfortunately, similar to the emergence of first-generation resistance mutations, acquired resistance develops in response to osimertinib and afatinib treatment [96]. The most common resistance following T790M is C797S. The replacement of C797 with serine ablates the covalent binding ability of these irreversible drugs and thus confers resistance in approximately 15–25% of patients treated with osimertinib [97, 98].

Uchibori et al. were able to identify a treatment option for T790M resistant mutant by simulating the triple mutant EGFR C797S/T790M/activating mutation and through structure-activity relationship analyses, brigatinib, a dual EGFR/ALK TKI, was proposed as a therapeutic agent [99]. Brigatinib binding to the C797/T790M/activating mutation EGFR ATP-binding site resembles that of the Echinoderm microtubule-associated protein-like 4 (EML4) -ALK (PDB: ID 6MX8). Kinase studies identified that the inhibitor is more potent against Del19/T790M/C797S than in L858R activating mutations. When screened against different cell lines presenting the triple mutants, brigatinib was the only drug to inhibit EGFR phosphorylation and its downstream signaling. The inhibitory effect of brigatinib on the triple mutant is improved once in combination with cetuximab, an anti-EGFR antibody already in use for cancer management [99].

L718 mutations significantly increase the IC50 value to osimertinib, with L718Q conferring the greatest resistance [100]. As previously mentioned, L718 is located on the P-loop in the proximity of the ATP-binding site and is important for the correct coordination of osimertinib during the covalent bond formation with C797. Substitution of leucine with glutamine sterically inhibits osimertinib binding due to the introduction of the larger, polar side chain, which decreases local hydrophobicity at the point of contact of osimertinib and spatially restricts its binding [100].

The L718Q mutant, in combination with L858R, confers resistance to gefitinib in
either the presence or absence of the T790M [100]. Following the substitution of leucine with glutamine, the local hydrophobicity is disrupted, which impairs gefitinib binding [99]. Interestingly, although the L718Q point-mutant confers resistance to osimertinib and gefitinib, a patient with advanced metastatic NSCLC harboring the L858R/L718Q double mutant was successfully treated with afatinib, indicating a furan moiety might be suitable in the presence of an L718 mutation [101].

L792F/Y/H mutations constitute approximately 1.5% of resistance to osimertinib, with the L792H conferring the most remarkable resistance [100]. Structural and mutagenesis analysis of the complex of WT EGFR/osimertinib (PDB: ID 4ZAU) showed that replacement of L792 with the aforementioned amino acids sterically inhibits the binding of osimertinib to the ATP binding cleft thus disrupting the correct orientation of the inhibitor and its pharmacological action [102, 103]. The G796C/D/R mutation has also been identified to confer resistance to osimertinib due to the steric hindrance effect imposed by the replacement of glycine to aspartic acid thus, impairing the formation of the previously mentioned hydrophobic sandwich [105].

A novel G724S point mutation, identified in a set of NSCLC patients and linked with resistance to osimertinib was studied using homology modeling [106]. Interestingly, the G724S only confers osimertinib resistance when combined with Del19 but not L858R mutants [106, 107]. Resistance arises through the complementary action of exon19del, which reduces β3-αC loop flexibility, and the G724S point mutation, which leads to the destabilization of the αC-in conformation [107].

EGFR mutations are rising faster than drug development can follow as seen by emerging clinical resistance to osimertinib and an associated poor prognostic for patients. Repurposing of already approved inhibitors can be of use as an accelerated methodology for clinicians as demonstrated for allopurinol and methotrexate, both initially developed to treat cancer but later repurposed for gout and rheumatoid arthritis [108]. The process of repurposing, despite being faster than following the pathway of developing a new molecular entity, remains hindered by the multitude of drugs to be assessed against a myriad of diseases, indicating a clear need for improvement on its methodological approach, opening a venue for application of in silico high throughput screening [109].

Molecular modeling techniques, as the ones previously described, can be used for a rapid assessment of novel mutations as shown by Kemper et al. with the triple mutation exon19del/T790M/P794L. A team of clinicians and structural biologists were faced with the emergence of a novel mutation, from a proline to a leucine at position 794, in addition to an exon 19 deletion and T790M. Despite the presence of the T790M, the patient was responding poorly to osimertinib. Through docking studies, the Molecular Tumor Board compared the drugs osimertinib and afatinib, providing insight into how afatinib could retain binding affinity, thus being a suitable therapeutic option [110]. This is an example that drug repurposing associated with the change of therapeutic indication of medication may also be used to reconsider drugs that were previously discarded.
1.4. Tyrosine-protein kinase ROS1

In cancer, protein kinases can be activated by mutations in nucleotides, gain or loss of chromosomes (somatic mutations or chromosomal alteration), and gene amplification. The discovery of the Echinoderm microtubule-associated protein-like 4 (EML4) and Anaplastic Lymphoma Kinase (ALK) fusion protein as a druggable target was a mark for treating lung adenocarcinomas. This finding highlights the importance of identifying and inhibiting such proteins to successfully treat a subset of NSCLC. Later, the proto-oncogene tyrosine-protein kinase ROS (ROS1) (UniProt ID: P08922), an orphan receptor tyrosine kinase, was associated with tumorigenesis and identified in an NSCLC patient with tumor progression. As ALK, ROS1 can be fused with different partners, the most common being the cell-surface receptor Cluster Differentiation (CD) 74 [111].

The fusion product conserves the ROS1 kinase domain which follows the conserved folding previously mentioned; an N- and C-terminal lobes connected by a hinge. Once again, the N-terminal contains the regulatory αC-helix and a beta-sheet composed of four strands. The αC-helix is connected to the β3 and β4 strands at its beginning and end, respectively. Strands β1 and β2 are connected through the G-loop that acts as a lid of the ATP binding pocket. In the active conformation, the activation loop is presented in an extended manner with the conserved Aspartic acid – Phenylalanine – Glycine (DFG) motif at its beginning. Once the protein assumes a DFG-in conformation, the DFG phenylalanine side chain is guided in a hydrophobic pocket and engages in hydrophobic stacking interactions [112]. This pocket contains residues from the αC-helix (L2000 and F2004), the DFG (F2103), and the other aromatic residues from the C-terminal (A2106 and F2075). A homologous pocket is observed on the ALK active conformation and is reported to contribute to the maintenance of the active state [113] (Figure 1.6).

The Protein Data Bank (PDB) presents, to this date, only four available structures of the ROS1 kinase domain (PDB: IDs 3ZBF, 4UXL, 7Z5W, 7Z5X), and all obtained through x-ray crystallography and co-crystallized with different ligands [114–116]. From those ligands, the multikinase inhibitor crizotinib was initially designed to inhibit the Hepatocyte Growth Factor Receptor (HGFR or c-MET) and later had its application extended to ALK and ROS1. When superimposing crizotinib complexes with ALK (PDB: ID 2XP2) ROS1 (PDB: ID 3ZBF) and c-MET (PDB: ID 2WGJ) it is possible to identify the R-enantiomer as the ligand in all structures. The drug binds to the ATP binding pocket in all proteins similarly, the 2-amino-pyridine ring engages in two hydrogen bonds with the backbone of ROS1 hinge residues, E2027, and M2029. Due to the R configuration, the methoxy unit containing the chiral carbon guides the 2,6-dichloro-3-fluorophenyl ring towards the DFG motif at the beginning of the activation loop (Figure 1.7) [114–116].

The major difference between all crystal structures concerns the pyrazole ring, while positioned between two key residues, L1951 and G2032 (ROS1), it retains a level of rotational freedom. Thus, the rotation of the pyrazole ring is associated with the piperidine ring positioning. The conformation observed for the pyrazole ring in the ROS1 complex is comparable to the position for the same drug on the ALK receptor (PDB: ID 2XP2), another target for this small molecule. The difference observed for c-MET can be credited to the presence of tyrosine (Y1159) in place of leucine as observed for ALK and ROS1 (L1198 and L2028, respectively) (Figure 1.7). The behavior of this five-member ring might affect the
binding, as interactions with the G-loop are pivotal for drug binding of type I inhibitors such as crizotinib [117]. Besides, Molecular Electrostatic Potential (MEP) calculations show a concentration of negative charges around the pyrazole and pyridine rings and positive charges in the substituted benzene ring which could explain the structure-activity relationship of crizotinib with cMET, ALK, and ROS1 targets [118].

Despite the success of crizotinib, there was a limitation on its ability to cross the blood-brain barrier and tackle metastatic incidences of ROS1+ tumors in the brain. For such, crizotinib was modified using a structure-based drug design (SBDD) approach leading to a macrocycle final product, lorlatinib. This macrocycle showed a lower propensity for p-glycoprotein (P-gp) efflux and improved BBB penetration [119]. The design of lorlatinib conserved the hinge-binding amino pyridine scaffold and the chiral center. The fluoro substituent in the aromatic ring was conserved while the two chlorines were not. The pyrazole ring was N-methylated and substituted at position 3 with a carbonitrile group. The pyrazole ring and the fluoro-phenyl ring are then connected by a methyl-amide motif leading to the cyclization of the structure [119, 120].

As expected, in the co-crystal with native ROS1 (PDB: ID 4UXL), lorlatinib binds in a comparable way to crizotinib. As observed for crizotinib, the aminopyridine core conserves the two hydrogen bonds to the kinase hinge. The fluoro atom polarizes the ortho aryl aromatic C-H groups leading to electrostatic complementarity with the backbone carbonyl at G2101, a glycine right before the DFG motif that has been explored for drug design [121]. In the crystal structure, the introduced N-methyl amide is close to the G-loop, interacting with the carbonyl group of L1951 and the backbone methylene of G1951. The amide carbonyl from the bridge group interacts with residues K1980 through
Figure 1.7 | (A) Overlap of crizotinib extracted from complexes with c-Met (yellow), ALK (magenta), and ROS1 (orange). Two-dimensional representation of first-generation inhibitors, (B) crizotinib, and (C) lorlatinib. Interactions were analyzed and plotted using Discovery Studio Visualizer and Pymol.

Despite the molecular similarity, lorlatinib and crizotinib presented different inhibition profiles even in the absence of mutations in the kinase domain. Comparison of the atomic displacement parameter, or B-Factor, for crystals structures of WT ROS1 in complex with crizotinib (PDB: ID 3ZBF) or lorlatinib (PDB: ID 4UXL), provided an overview of how the presence of the ligands can influence the flexibility of key motifs. While for both
structures the highest values for the B-factor were associated with the G-loop and the activation loop, these regions presented a better resolution for the 4UXL crystal structure. The fact of the G-loop region being unresolved for crizotinib indicates a higher flexibility of this motif in the presence of the drug. Overall, analysis of the normalized B-factor of both structures indicated that lorlatinib stabilizes the ROS1 kinase domain more than crizotinib. This effect is especially observed for residues L1951, V1959, and V1979, all located in the N-terminal β-sheet [122].

Approaches relying on docking are often used to identify novel scaffolds, however, they present a low success rate [123]. The low rate of success can be associated with the onerous step of identifying hits within an extensive library, which can be time-consuming and often depends on visual analysis. Altogether, virtual screening frequently provides false positives that do not go further in the development pipeline. Aiming to improve the application of docking in drug discovery, AstraZeneca (AZ) proposed a combination of Structure and Ligand Based Drug Design, SBDD, and LBDD respectively. Starting from a set of five molecules identified by different approaches but with considerable ROS1 potency and lack thereof over neurotrophic receptor tyrosine kinase 1 (TrKA), a kinase anti-target, AZ identified novel ROS1 inhibitor prototypes. Besides consolidating the efficacy of the use of the FastROCS approach and its use in a cloud computing platform, this work also provided two crystal structures of novel ROS1 and ligand complexes [116, 124].

One of the novel structures shows the binding of an Insulin-like Growth Factor Receptor (IGF-1R) inhibitor (ligand 1) to the ROS1 kinase domain (PDB: ID 7Z5W). In this crystallized structure, the hinge binding motif is the 2-aminothiazole group that engages in two hydrogen bonds with the backbone of M2029 through both nitrogen atoms, the aromatic one acting as an acceptor and the other as a donor. Another hydrogen bond is mediated by a water molecule between the pyrimidine group and the side chain of D2033. This ring is also positioned between residues L1951 and G2032, comparable to the pyrazole ring from crizotinib. The isoxazole and the pyrrolidine rings contribute as linkers to guide the 3-methyl pyrazine ring toward the beginning of the activation loop. In one of the monomers that compose the crystal structure, it is possible to identify an approximation of this motif to the side chain of the D2102 of the DFG motif (Figure 1.8) [116].

In parallel, Petrovic also crystallized an ALK inhibitor derived from crizotinib with the ROS1 kinase domain (PDB: ID 7Z5X). This analog conserved the hinge binding amino pyridine scaffold while modifying the two tails to improve the pharmacokinetic profile and inhibitory profile. The piperidine ring was substituted by a 1,2-propanediol group which interacted with the residue D2033, and a network of hydrogen bonds mediated by water molecules. Despite conserving the fluor atom from the 2-6-dichloro-3-fluoro phenyl ring, one chlorine atom was removed while the other was substituted by a triazole ring. In this crystal structure, the five-membered ring engages in an intramolecular hydrogen bond with the hydroxyl attached to the thiazole ring through a water molecule bridge (Figure 1.8) [116].

Other ALK inhibitors also showed effective inhibition of ROS1. Entrectinib, a 3-aminoindazole derivative was initially designed as a type I ALK inhibitor but with comparable sensitivity for ROS1 [125]. Although a crystal structure of ALK in complex with
entrectinib is available, ROS1 lacks an experimentally determined complex structure. However, a study using MD simulations and Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) calculations proposes a mechanism for the binding profile of entrectinib. Overall, the binding affinities obtained from MMPBSA calculations of this drug to ALK and ROS1 are comparable, and both rely on similar residues within the ATP binding pocket. In addition, the major role of L2086 is featured for the stabilization of the ROS1-entrectinib complex due to its consistent interaction with the 3,5-difluoro benzyl ring (Figure 1.9) [125].

Through homology modeling, using a DFG-out inactive ALK structure (PDB: ID 4FNY), Davare et al were able to model the inactive conformation of the ROS1 kinase domain [113]. This model showed the absence of the phenylalanine cluster that gave place to a new pocket, the selectivity pocket. The selectivity pocket is known to be present in the inactive kinases domain and provides a less conserved pocket that can be explored for selectivity in the kinase drug design [35]. This pocket is located between the αC-helix and the catalytic activation loop with the latter being dislocated towards the G-loop. Despite being modeled after an ALK inactive structure, the selective pocket for ROS1 is considerably different from the ALK specificity pocket due to the nature of the amino acids present. In this model, the inactive conformation corresponds to the phenylalanine of the DFG motif occupying the ATP binding pocket (DFG-out) and is a suitable target for type II inhibitors such as cabozantinib or foretinib (Figure 1.10) [113].
Figure 1.9 | Interaction map for entrectinib in the ROS1 active proposed by Menichincheri et al. [125].

Analysis of MD simulation of the inactive ROS1 WT indicated a more rigid P/G-loop. The same pattern is observed for the activation motif, in the inactive ROS1 model this motif is more rigid due to its proximity to the P-loop and the αC-helix. The rigidity observed for the αC-helix is due to the proximity of the Q2012 residue to the C-terminal of the same helix, restricting its movement toward the catalytic site which can explain the average pocket volume for the ROS1 and ALK, being 186 Å³ and 235 Å³, respectively. Altogether, it was demonstrated that despite the significant sequence homology between ROS1 and ALK, discrepancies in the rigidity and motion of regulatory motifs contribute to the difference observed in the specificity pocket of each kinase and their pockets’ volumes [113].

Furthermore, the same inactive ROS1 model was used in Relaxed Complex Scheme (RCS) studies to propose a binding pose for cabozantinib, a type II ROS1 inhibitor. The proposed binding mode shows the quinoline moiety interacting with the hinge through hydrogen bonds with residue E2027 and M2029, as crizotinib. The aryl linker engages in aromatic stacking with residue F2103 from the DFG motif while the carboxamide moiety and residue K1980 interact through a hydrogen bond. This fluorophenyl tail is proposed to be buried in the specificity pocket and interacts with F2004, a residue present in the phenylalanine pocket located at the end of the αC-helix, and F2075 through aromatic stacking interaction. Additional residues M2001, L2070, and L2100 contribute to hydrophobic interactions with this drug (Figure 1.11) [113].

The same study also proposed a binding pose for foretinib, a potent ROS1 inhibitor structurally related to cabozantinib [126]. The predominant pose obtained from the RCS approach using the apo MD simulations is like the one proposed for cabozantinib with
the additional morpholine ring interacting with residues K1967 and E2030. A closely related binding conformation was also observed for the same molecule but with the cMET receptor (PDB: ID 6SD9) [127]. However, a minor cluster of poses obtained from the docking study proposed a reversed position of the drug in the binding pocket, with the morpholine closer to the αC-helix. This alternative binding mode proposes a second interaction mode and the reason why this mutation retains sensitivity to foretinib even when cabozantinib does not [113].

Gatekeeper residues are in the hinge motif’s N-terminal end and are responsible for the accessibility to the “back pocket” of kinases. A point mutation in this position is the most common resistance mutation to kinase inhibitors and is described for many different kinases [128]. In ROS1, the gatekeeper residue is the L2026 and the most common mutation is the mutation of the leucine into a methionine (Figure 1.12) [129].

Lorlatinib effectively inhibited Ba/F3 cells harboring the CD74-ROS1\textsuperscript{L2026M} fusion kinase [115]. While this mutation decreases crizotinib potency, lorlatinib was efficient in inhibiting the growth of cells expressing CD74-ROS1\textsuperscript{L2026M} in a dose-dependent manner [115]. Despite not having an experimentally determined structure for this mutant, earlier studies overlap the native ROS1 kinase domain to the analogous ALK\textsuperscript{L1196M} crystal structure and indicated that the presence of a methionine at position 2026 in the ROS1 protein would likely be accommodated due to lorlatinib lacking the piperidine substituent [115].

Once analyzed under the perspective of a combination of MD and free energy calculations, the resistance to crizotinib shows the conservation of the hydrogen bond between the aminopyridine ring and residues E2027 and M2029. However, the contribution of both hinge residues for the mutant was decreased when compared with the WT. Although a meager positive impact of the mutated M2026 in the binding affinity was described, the reduced binding affinity of crizotinib was credited to increased entropic terms [129].
As observed for the ALK fusions, the treatment of crizotinib in ROS1-positive patients also led to the rise of point mutations in the kinase domain. The first mutation detected was in a lung cancer patient initially treated with crizotinib in a residue exposed to the solvent. The authors reported the mutation G2032R as resistant to crizotinib (Figure 1.12) [114]. This mutation is analogous to the G1202R ALK mutation that is also linked with the crizotinib resistance [130, 131].

Despite having a determined crystal structure of the wild-type ROS1 kinase domain complex with crizotinib, attempts to obtain the mutant G2032R mutant were unsuccessful. Thus, resourcing to homology modeling to analyze the impact of this mutation on the ATP binding pocket was performed. Analysis of the G2032R model indicates that the resistance to crizotinib in the mutant would be caused by the side chain of the R2032 crashing with the piperidine ring from crizotinib. It was also proposed that the mutation would not hamper ATP binding, as the ATP concentration required to achieve half-maximal enzyme velocity was reduced for this mutant compared to the non-mutated ROS1 [114].
Figure 1.12 | Representation of point mutations in the ROS1 kinase domain: (A) L2026M, (B) G2032R, and (C) D2033N. In all figures, the native ROS1 is represented in green color with the mutated residues represented in yellow, while superimposed with the native counter partner residue. Key structural motifs such as the G-loop and the αC-helix are highlighted in the figures to aid identification of location

Later, free energy calculations and MD simulations of the apo ROS1 and the ROS1-crizotinib complex provided a more in-depth interpretation of the effect of this mutation. The analysis of a dihedral defined by the Cα carbon of the residues S1953-A1955-E1958-R2083 showed higher flexibility of the G-loop in the presence of the G2032R mutant than for the WT ROS1 thus indicating a more opened structure of the G-loop in the apo state. Although a similar behavior in the G-loop is observed in the crizotinib-bounded state with a smaller difference between the mutant and nonmutated kinase domain, the comparison of the apo WT and the bounded G2032R dihedral distribution was rather similar. The opened conformation of the G-loop in the mutant was attributed to the inserted arginine, which served as a scaffold for lifting the loop. Despite bringing remarkable insight into the position of the G-loop, the use of classical MD failed to explain the resistance of the mutant to crizotinib [132].

In the same study, using funnel-based well-tempered metadynamics and umbrella sampling to calculate the absolute binding free energy of crizotinib, it was possible to support the correlation between the P-loop rigidity and the effect on the drug binding. The results show that the G2032R mutants sustain a more open conformation of the G-loop and impact the interaction between the residue L1951 and the drug. This residue in the complex WT ROS1 and crizotinib was responsible for the highest contribution to the drug-binding [132]. Another study using MD simulations of the apo WT and ROS1\textsuperscript{G2032R} supports the overall protein rigidity found in the mutant, especially for the G-loop and the A-loop. The different mobility observed between native and mutant ROS1 was credited to the loss and gain of intramolecular interactions. A stabilizing salt bridge was identified between the guanidinium group of R2032 and the carboxyl from E1961 in the mutant. Meanwhile, the stabilizing hydrogen bond between E1961 and R1948 identified in the apo WT ROS1 simulations was lost in the presence of the mutation. This study showed that not only the mutation of the glycine to an arginine leads to a steric hindrance to crizotinib binding but also how the G2032R mutation modifies the intramolecular interactions in the kinase domain that could account for drug resistance [113].

Supporting these findings, MMPBSA calculations showed a decrease in the enthalpic term of free energy calculations. The decrease was especially associated with residues E2027 and M2029, both involved in hydrogen bonds with the aminopyridine scaffold [129]. Since ROS1 G2032R was proved resistant to crizotinib, it was necessary to seek new
molecular entities able to inhibit this variant.

In the same study, cabozantinib, an analog to foretinib showed effective inhibition of the CD74-ROS1\(^{G2032R}\) cell line. The docking of cabozantinib in the ensemble for either the native and the mutated ROS1 was considered favorable and mirrored the in vitro assessment [113]. Despite the limited data, the aforementioned description of a probable foretinib binding mode to ROS1 can explain the sustained sensitivity of the G2032R mutant to this type II inhibitor [126].

Despite the G2032R mutation association with crizotinib resistance, its analog lorlatinib was efficient/ maintained activity in the inhibition of the catalytic activity of BaF3 cells harboring the fusion CD74-ROS1\(^{G2032R}\) [115]. As previously mentioned, the use of homology modeling implicated resistance of this mutation to crizotinib to a clash between the side chain of the mutated R2032 and the pyrazole and piperidine ring. By using the same approach, it was proposed that the clash would be significantly reduced due to the absence of the piperidine ring and the conformational restrain of the pyrazole ring [115].

This hypothesis was later supported by MD and free energy calculations. Despite the MM-PBSA/GBSA calculations differing from experimentally determined binding free energies, the magnitude difference observed between the WT and G2032R was replicated. In the mutant, interaction with residues E2030, M2029, and L1951 was weaker, comparable with a similar analysis for crizotinib in the WT binding pocket. In addition, it was shown that despite the mutated R2032 side chain favorable interaction with lorlatinib, its contribution was comparable to the G2032 in the WT protein. As disclosed to crizotinib, the reduced affinity of lorlatinib was associated with an increase in the conformational entropy [133].

The aspartic acid at position 2033 is located at the C-terminal of the hinge (Figure 1.12). The ROS1\(^{D2033N}\) was initially found in a patient with resistance to crizotinib which was later confirmed through Ba/F3 cells expressing CD74-ROS1\(^{D2033N}\) [134]. Cabozantinib was then administered to the patient who presented a positive response. The MD simulations and docking studies combined highlighted the strong electrostatic interaction between the piperidine ring of crizotinib and the negatively charged side chain of the D2033. However, the mutant loses this interaction as the inserted asparagine lacks the negatively charged side chain. Instead, the N2033 side chain reorients towards the G loop and engages in a water-mediated hydrogen bond with L1951, leading to a slight loss of flexibility in the G loop [134].

The MD simulations also indicated that the presence of asparagine at position 2033 induced neighboring residue reorientation that would clash with the protonated piperidine ring of crizotinib. Structural alignment of the structural ensemble obtained from MD simulations with the crystal complex of ROS1-crizotinib indicated an electrostatic repulsion between the charge of the positively charged nitrogen from crizotinib and N2033. This is not observed for cabozantinib, as the nearest atom of cabozantinib was 5 Å from position 2033 in both native and mutant [134].

The activation loop is not fully characterized in all available crystal structures. This is understandable given the flexible nature of this loop. Interestingly, the exposition of BA/F3 cells expressing native CD74-ROS1 to high concentrations of cabozantinib and foretinib led to the emergence of mutation D2113N as resistant to both type II inhibitors.
In the inactive ROS1 model used by Davarre, the aspartic acid at position 2113 interacts with the R2116 while being repelled by the E2120 at the A-loop [113].

The mutation D2113N was initially described in a cell-based resistance screening. This mutation was linked to resistance to type II inhibitors while retaining sensitivity to ROS1/ALK dual inhibitors such as crizotinib, ceritinib, and brigatinib. MD simulations of the inactive conformation of ROS1, native and mutated, indicate the stiffness of the A-loop in the presence of the mutation as the culprit for resistance to type II inhibitors. Besides, the replacement of the aspartic acid with the asparagine side chain nullifies the repulsion from the E2120 residue while guiding this residue into a hydrogen bond with the mutated side chain [113].

As previously mentioned, simulations of the native ROS1 show high occupancy for a salt bridge between residues E1997 (αC-helix) and R2107 (A loop), an interaction also described for other kinases in their inactive states [135]. While the ROS1WT lacked such interaction in the active simulations, the inactive simulations showed a consistent presence of this salt bridge. However, for the D2113N mutant, the frequency of this interaction is diminished. Analysis of the simulations for the inactive conformation of the ROS1D2133N mutant suggests a displacement of the R2107 residue by the mutated sidechain leading to a loss of the salt bridge. This leads to the E1997 side-chain reorientation towards the specificity pocket, partially occupying it and possibly crashing with a type II inhibitor. Interestingly, docking studies mimic the in vitro and MD simulations, showing crizotinib as a more favorable inhibitor than cabozantinib and foretinib [113].

Cell-based resistance screens performed with Ba/F3 performed in either native or G2032R CD74-ROS1 also provided additional identification of single or double mutants. Once performed on native CD74-ROS1 cell lines, such studies linked resistance to cabozantinib with mutations involving F2004 and F2075, residues described as interaction spots for type II inhibitors [113]. Previous reports already indicated such residues as possible hotspots for type II resistance due to analogous positions on ALK resistance to kinase inhibitors [136].

Additional point mutations have been described for ROS1 but often with different fusion partners. A more extensive revision on single mutations and their associated fusion partner can be read in other papers [137, 138]. It is interesting to highlight that most of the in silico studies focus on the classical G2032R and L206M thus leaving a gap for other mutations and their impact in the ROS1 kinase domain and further mechanism of resistance.

Despite compound mutations being commonly found in most of the kinases targeted in clinical studies, up to the moment we finalized this review, there is no report of double ROS1 mutations found in patient-related material. Resistance screening exposing cells harboring ROS1 fusion attempts to predict future double mutations and their response to kinase inhibitors. Treatment of Ba/F3 cells presenting CD74-ROS1G2032R with increasing doses of cabozantinib or foretinib led to the rise of secondary mutations at positions E1974, F2004, I2009, E2020, F2075, N2112, D2113, R2116, M2128, D2143, L2223 and N2224 (Figure 1.13) [113]. However, the structural impact of these double mutants on ROS1 dynamics and their limited sensitivity to kinase inhibitors remains uncharted.

The disclosure of the role of ROS1 fusion proteins in cancer provided a novel druggable target. The high similarity between ROS1 and ALK’s ATP binding pocket made conceivable
the extension of ALK inhibitors use to ROS1 inhibition. The co-crystallization of crizotinib with the ROS1 kinase domain drew a parallel on the binding mode of this small molecule and how comparable is to the same drug in other kinases. Despite the availability of other crystal structures bearing different ligands, there are no experimentally determined structures of ROS1 mutants.

Mutations in the ROS1 kinase domain are linked with resistance to type I inhibitors and were primarily found in patients treated with crizotinib. The G2032R mutation is in the area exposed to the solvent and has been extensively characterized as a crizotinib-resistant mutation. Molecular modeling of this mutant through different methods shows a clear impact of the mutant in the overall flexibility of the kinase domain besides possible limitation on crizotinib binding due to its substitutions. Additionally, such an approach explains why despite ROS1\(^{G2032R}\) being resistant to crizotinib, it sustains sensitivity to lorlatinib, a crizotinib structural analog.

Gatekeeper mutations are a common class of mutation in kinases. This mutation often renders kinase inhibitors ineffective due to an increased affinity to ATP of steric clashes. Reports of crizotinib interaction with ROS1\(^{L2026M}\) associated with the resistance to entropic changes that were minimized for lorlatinib. Yet, a deeper knowledge of how this mutation affects the kinase domain of ROS1 and the efficacy of other kinase
inhibitors is needed. The valuable characterization of the native and mutated (D2113N) inactive conformation of ROS1 provides insight into the binding of type II inhibitors. The contribution of these models and further MD simulations become more remarkable due to the lack of an experimentally determined structure of the ROS1 kinase domain in its inactive conformation.

In summary, there is an imperative need to characterize ROS1 and its mutations either by experimental or in silico techniques. Despite the indisputable relevance of experimental data, computational tools have shown their potential to bridge gaps when atomistic details are needed.

1.5. GOALS AND OUTLINE OF THIS THESIS

The dynamic interplay of phosphorylation and dephosphorylation plays a pivotal role in regulating crucial biological processes within cells. However, disruptions in these events are closely linked to the development of cancer. In malignant cells, such disruptions are frequently attributed to mutations in kinases, enzymes responsible for the phosphorylation of substrates, and the subsequent activation of downstream pathways. These mutations confer a proliferative advantage to cancer cells by driving hyperactivation of the kinase, resulting in sustained catalytic activity.

The mutational landscape of kinases is vast, despite the limited possibilities of point mutations; the prospect of insertions and deletions of residues expands this landscape to an unmeasurable extent. Although not all mutations might affect the kinase structure or activity (i.e. passenger mutations), there are many with a remarkable impact. Despite the extreme success of medicinal chemistry in tackling these mutated kinases with the development of potent kinase inhibitors or antibodies, the emergence of oncogenic mutations is in a fast lane that cannot be stopped.

Assessing every possible mutation experimentally in just one kinase would demand not only a humongous amount of time but also intense financial input. And even with all the time and money, these experimental models might fail; the cell line containing the desired mutation might not survive or the recombinant expression of the mutated kinase domain might not yield sufficient material to enable subsequent experiments. From this perspective, molecular modeling can provide a reliable way of simulating these mutated systems and their interactions with other molecules without the need for extensive experimental data as input.

In Chapter 1, we presented a comprehensive review of studies that employ in silico methods to analyze mutations in the EGFR and ROS1 kinase domains. This review demonstrates how these techniques, either leveraging prior experimental data or conducted in conjunction with wetlab investigations, can yield invaluable insights into the mechanisms underlying the activation of these kinase domains.

In Chapter 2, molecular dynamics simulations of the active and inactive conformations of the native EGFR kinase domain shed light on possible similarities and differences that can be used as parameters of comparison upon mutation. Less common activating point mutations on exon 20 are also studied and their impact on the conformational landscape and dimerization process. In this chapter, we propose a mechanism of activation for mutations S768I, D770N, and R776G.
Mutations on exon 20 can also lead to the insertion or deletion of residues within the $\alpha$C-$\beta$4 loop on the EGFR kinase domain. In Chapter 3, we simulate four mutations that lead to insertions varying from one to three residues in different positions of the $\alpha$C-$\beta$4 loop: D770-N771insG, D770-N771insGF, D770-N771insGTD, and N771-H773insNPH. To understand the impact of these mutations on the active-inactive balance of the EGFR kinase domain we model and simulated the active and inactive conformation of the native species and all the variants.

Chromosomal rearrangements can fuse the kinase domain of ROS1 with different partner proteins, making this kinase an interesting target in cancer. In Chapter 4 we focus on point mutations at the ROS1 kinase domain that, despite being oncogenic and associated with resistance to kinase inhibitors, are not located directly in the catalytic site. We investigate mutations L1982F, S986F, and S1986Y, located in the $\beta$3-$\alpha$C loop, and their impact on the regulatory $\alpha$C-helix and consequently on the ATP binding pocket.

It is common for the emergence of secondary mutations upon treatment with kinase inhibitors. In Chapter 5, an N-Ethyl-Nitrosourea (ENU) mutagenesis screening showed the emergence of a ROS1 double mutant G2032R-L2026M resistant to two kinase inhibitors, crizotinib, and lorlatinib. In this chapter, we combined molecular dynamics simulations and molecular docking to explain the mechanism of resistance of the double mutant to crizotinib and lorlatinib.

And finally, in Chapter 6 through simulations of a model of the inactive ROS1 kinase domain, we draw a comparison between the active and inactive conformation of the ROS1 kinase domain. Additionally, we explore the effect of mutations F2004C and F2004V in both states.

REFERENCES


[64] K. Li, M. Yang, N. Liang, and S. Li, Determining egfr-tki sensitivity of g719x and other uncommon egfr mutations in non-small cell lung cancer: Perplexity and solution, Oncology reports 37, 1347 (2017).


[70] X. Ding, X. Liu, X. Song, and J. Yao, Chemotherapy drug response to the l858r-induced conformational change of egfr activation loop in lung cancer, Molecular informatics 35, 529 (2016).


1


[78] M. Bello, Binding mechanism of kinase inhibitors to egfr and t790m, l858r and l858r/t790m mutants through structural and energetic analysis, International journal of biological macromolecules 118, 1948 (2018).


C.-H. Yun, K. E. Mengwasser, A. V. Toms, M. S. Woo, H. Greulich, K.-K. Wong, M. Meyerson, and M. J. Eck, The t790m mutation in egfr kinase causes drug resistance by increasing the affinity for atp, Proceedings of the National Academy of Sciences 105, 2070 (2008).


A. Masood, R. K. Kancha, and J. Subramanian, Epidermal growth factor receptor (egfr) tyrosine kinase inhibitors in non-small cell lung cancer harboring uncommon


